MODELING OF PDFS OF NON-GAUSSIAN DATA AND APPLICATION TO WIND PRESSURE COEFFICIENTS ON LOW-RISE BUILDING

By

LUPING YANG

A THESIS PRESENTED TO THE GRADUATE SCHOOL
OF THE UNIVERSITY OF FLORIDA IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE

UNIVERSITY OF FLORIDA

2011

To my parents Qingcheng Yang and Chaoqun Pang

ACKNOWLEDGMENTS

First of all, I wish to express my gratitude to my committee chair David O. Prevatt. This work would not have been accomplished without his patient guidance, continual support and great encouragement. I am also grateful to my co-chair and research advisor Dr. Kurtis R. Gurley. Working with Dr. Gurley has been a great pleasure for me. His insightful knowledge and amiable smile always gives me inspiration and encouragement when I encounter difficulties. His guidance to me has not only been limited to study and research but also extended to the way of thinking, the manner of interacting with people and the attitude of life. I wish to express my sincere gratitude to Dr. Sergei V. Shabanov in the Mathematical department for serving on my advisory committee.

I would like to thank my family, Mr. and Mrs. Yang and my brother Wei Yang, for all their love, care and support. They have made me feel happy and have given me faith to study from thousands of miles away. Lastly, I wish to express my gratitude to my friends, colleagues and the hurricane research group at the University of Florida. Studying and working at UF has been a valuable and pleasant experience in my life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Abstract of Thesis Presented to the Graduate School
of the University of Florida in Partial Fulfillment of the
Requirements for the Degree of Master of Science

MODELING OF PDFS OF NON-GAUSSIAN DATA AND APPLICATION TO WIND PRESSURE COEFFICIENTS ON LOW-RISE BUILDING

By

Luping Yang

December 2011

Chair: David O. Prevatt
Cochair: Kurtis R. Gurley
Major: Civil Engineering

During their service life, structures are often subjected to severe loads, whose probabilistic distribution functions (PDFs) significantly deviate from Gaussian distribution significantly, especially in the extreme regions. To assure the safety and reliability of structures, it is necessary better represent the PDFs of these severe loads in the tail regions. In the specific field of wind engineering, hurricane damages to light framed wood structures (LFWS) account for a disproportionately large proportion of economic losses. During these extreme events, the distribution of wind loads can differ dramatically from Gaussian distribution. To improve the prediction of extreme loads and fatigue accumulation and further reduce the economic loss caused by hurricanes, an accurate representation of PDFs of the wind loads in the tail regions is required.

In this research, three PDF models are studied to generate the Probability Density Function of Non-Gaussian data. They are Pade-Laplace Method (PLM), Maximum Entropy Method (MEM) and Hermite Polynomial Method (HPM). Firstly, each of the three methods is formulated or improved, and then applied to modeling the PDFs of experimental data. In this process, it is found that Pade-Laplace method generate

negative PDF value and is thus not admissible for modeling PDFs of Non-Gaussian data. Then, the reason why PLM generates negative value is studied. For Maximun Entropy Method, oscillating tail behavior of MEM is observed and alternative constrains are applied to overcome the oscillating tail behavior of the traditional MEM, but MEM based on alternative constrains are found to be not robust. HPM is improved by establishing an accurate relationship (one set of nonlinear equations) between the parameters and higher moments, i.e. skewness and kurtosis. By this mathematical relationship, an effective region of HPM is obtained, which is critical for the application of Hermite model. Also the set of nonlinear equations are approximated using surface fitting algorithm to simplify the application of HPM.

After the formulation and examination of the PDF models, they are applied to wind pressure data, together with common PDF models, such as Gaussian, Lognormal, Gamma distribution, etc. In this process, wind tunnel data on a 1:50 scale gable roof house and full-scale data was analyzed to investigate the spatial variation of Non-Gaussian characteristics of wind pressure coefficients. This analysis shows how the patterns of spatial variation of skewness and kurtosis change with the region of the roof and flow characteristics. Also the uncertainty associated with skewness and kurtosis, is analyzed to justify the necessity to use Maximum Likelihood Method (MLM), in roder to reduce the uncertainty in data statistics. After that, test of Goodness of Fit (GOF) are conducted among the proposed PDF model and common PDF models to demonstrate the flexibility and robustness of proposed Hermite PDF model. Finally, to show the implication of the excellent performance of HPM, the probability of non-exceedance

predicted by HPM and other common PDF models are compared. The results show that HPM has greatly improved the accuracy of the probability of non-exceedance.

To sum it up, It is found that HPM accurately represents the probability content of both weak and strongly non-Gaussian data, whereas the more common models have mixed performance, and provide poor results for the strongly non-Gaussian data. HPM has high potential in aiding the modeling of extreme wind pressures on bluff bodies.

CHAPTER 1
INTRODUCTION

Motivation

The methodology of probabilistic design (i.e. Load Resistance Factored Design) has long been adopted. When using a probabilistic approach to design, the designer no longer thinks of each variable as a single value or number. Instead, each variable is viewed as a probabilistic distribution. In structural safety assessment, structures are designed to withstand severe loads during extreme events, such as earthquakes and hurricanes, during which the distributions of loads deviate greatly from Gaussian. According to reliability theory, the probability of failure highly depends on the distribution in the tail regions, which means generating accurate PDFs of extreme loads is essential for determining the safety of structures.

In wind engineering, the probabilistic modeling of wind pressure is critical for extreme value estimation and fatigue analysis. It was found that non-linear relationships between pressure and velocity result in probability distributions of wind pressure which depart significantly from Gaussian, particularly in their tail regions(Holmes 1981). Since prediction of extreme values of wind pressure depends on its parent distribution, deviations of parent distribution from Gaussian distribution will have a significant impact on the predicted extreme values. This non-Gaussian effect can increase the rate of fatigue damage accumulation (Kumar and Stathopoulos 1998).

For these two reasons, it is important to establish an appropriate probabilistic distribution model for non-Gaussian data and structural responses, which can generate a good probabilistic representation of the loads with an emphasis in the tail regions.

To find probabilistic representations of loads, the common practice is to fit a set of candidate families of common PDF models to data and then choose the best model from among the set of candidate families of distributions. However, common PDF models have been demonstrated to have a limited capability to capture the tail distribution of non-Gaussian data.

The PDFs of wind pressure over a building surfaces vary dramatically from region to region, with some distributions being Gaussian in nature while others (mainly along the edges) being highly non-Gaussian distributions, and so a flexible and robust PDF model is required to represent the wind pressure distributions over this range of PDFs (e.g.(Sadek and Simiu 2002), (Cope et al. 2005)). However, typical PDF models are not flexible enough to model the PDFs with widely varying statistical properties.

Therefore, the motivation of this thesis is to develop, improve and compare the methods of modeling and use of PDFs for non-Gaussian and Gaussian data. The goal is to identify an appropriate universal PDF model that can provide consistently accurate probabilistic representations of all wind pressure distributions, particularly in the tail regions regardless of where on structure's surface the pressures are being investigated.

Objective

The three main objectives of this investigation are to:

1. Analyze wind pressure data to understand the variation of statistical properties of wind pressure and quantify the uncertainty in these statistics;
2. Develop and improve non-standard PDF models for the application to modeling Non-Gaussian data;
3. Identify an appropriate robust PDF model for probabilistic representation of wind pressure.

Organization

Chapter 2 provides general background for modeling PDFs of data, including statistical properties of random variables, common PDF models, and PDF models based on higher-order moments.

In Chapter 3, wind tunnel test and for-scale test, from which wind pressure coefficients were obtained, are introduced. Then statistical analysis is conducted to show the spatial variation of Non-Gaussian characteristics of wind pressure coefficients. Then the variation of the statistics of wind pressure is analyzed to demonstrate the uncertainty embedded in the statistics.

The Pade-Laplace method is in discussed in Chapter 4. Firstly, formulation of the method is introduced and then the method is applied to model theoretical PDFs and experimental data. However, the PDFs generated by PLM are found to have negative value, which limits the application of PLM. And the reason resulting in the negative value is studied.

In chapter 5, Maximum Entropy Method is studied. The main focus of this chapter is to use alternative constraints to eliminate the multimodal of PDFs generated by MEM. However, the MEM based on alternative is not robust for application.

Chapter 6 presents Hermite Polynomial Method. Two main improvements are made to HPM: the effective region of HPM is established to make the method complete; the nonlinear relationship between c_3 , c_4 and skewness and kurtosis is fitted to aid practical application. The method maximum likelihood is applied to HPM to estimate the input parameter to reduce the uncertainty embedded in statistics and improve the accuracy of HPM.

In chapter 7, the data analysis is presented. The pairs of skewness and kurtosis are compared with effective region to show the applicability of HPM. Then comparisons of the goodness of fit of the common PDF models and HPM PDF models are performed. Also the PDFs models are applied to estimated the probability of non-exceedance in the left tail region.

In chapter 8, a summary of all the work is included. The conclusion of the study and recommendation are given.

Chapter 2
LITERATURE REVIEW

Statistical Properties of Random Variables

In probability and statistics, moments, e.g. mean value and variance, are a quantitative measure of the distribution of a random variable. If all moments of a random variable are given, its probability density function could be completely determined. In practice, however, it is not feasible to obtain all the moments of a random variable, even if they actually exist. In other words, the distribution of a random variable is typically approximated in terms of certain descriptors of the random variable. The most important of these are the moments of the random variable. If the form of the distribution function is known, these moments can be used to determine the shape parameters, such as the mean and standard deviation of Gaussian distribution, of the distribution function.

Next, the first four central moments, which are frequently discussed and used in this thesis, are introduced.

Mean (Expected Value)

The mean (expected value) of a random variable is the weighted average of all possible values that this random variable can take on. The weights used in computing this average correspond to the probabilities in case of a discrete random variable, or densities in case of a continuous random variable, respectively,

$$E(X) = \sum_{i=1}^{n} p_X(x_i) x_i \tag{1-1}$$

$$E(X) = \int_{-\infty}^{+\infty} x \cdot p_X(x) dx \tag{1-2}$$

where, p_X is the mass of a discrete random variable or the density of a continuous random variable.

From Figure 2-1, mean value indicates the position of a distribution, if the mean changes, the corresponding distribution will translate along the abscissa.

Variance and Standard Deviation

The variance is used as a measure of dispersion or variability of a random variable; that is, the quantity that gives a measure of how closely the values of the variant are clustered about the central value.

If the deviations are taken with respect to the mean value, then a suitable average measure of dispersion is variance. For a discrete random variable X, the variance of X is

$$Var(X) = \sum_i (x_i - \mu_X)^2 \cdot p_X(x_i) \tag{1-3}$$

in which $\mu_X \equiv E(X)$.

If X is continuous, the variance is

$$Var(X) = \int_{-\infty}^{+\infty} (x - \mu_X)^2 \cdot p_X(x) dx \tag{1-4}$$

A more convenient measure of dispersion is the square root of the variance, or the standard deviation, σ; that is

$$\sigma_X = \sqrt{Var(X)} \tag{1-5}$$

As shown in Figure 2-2, it is observed that larger Standard Deviation indicates that the data is more deviated from the mean.

Skewness

skewness is a measure of the asymmetry of the probability distribution of a random variable. It is defined as:

$$\gamma_3 = \frac{E^3(X - \mu_X)}{\sigma_X{}^3} \tag{1-6}$$

The skewness value can be positive or negative, or even undefined. Qualitatively, a negative skewness indicates that the tail on the left side of the probability density function is longer than the right side and the bulk of the values (including the median) lie to the right of the mean (shown in Figure 2-3).

From Figure 2-3, a positive skew indicates that the tail on the right side is longer than the left side and the bulk of the values lie to the left of the mean. A zero value indicates that the values are relatively evenly distributed on both sides of the mean, typically but not necessarily implying a symmetric distribution.

Kurtosis

Kurtosis is a measure of the "peakedness" of the probability distribution of a real-valued random variable, or its symmetric deviation from the Gaussian distribution.

$$\gamma_4 = \frac{E^4(X - \mu_X)}{\sigma_X{}^4} \tag{1-7}$$

Higher kurtosis means more of the variance is the result of infrequent extreme deviations, as opposed to frequent modestly sized deviations. As shown in Figure 2-4, larger kurtosis means the distribution is more peaked and has fatter tails.

Based on above discussion, we can see that statistical properties, i.e. the mentioned moments, can be used to discern a great deal about a variable. With these

limited information about the first four moments of a random variable, reasonably accurate PDF models could be better established for a dataset.

Common PDF Models

Due to the temporal and spatial variations of wind speed and wind-induced forces on a building a statistical approach serves as a good tool to deal with the random characteristics of wind. In this context, identifying or estimating the PDFs of wind loads is one of priorities for reliability analysis and extreme value analysis and has been studied by a number of researchers ((Holmes 1981), (Li, Calderone and Melbourne 1999) and (Cope et al. 2005)). The common way of modeling the PDFs of wind loads is:

1) Calculate the statistical properties, the moments, from field or experimental data;
2) Fit existing PDFs models to collected data using these statistical properties;
3) Evaluate the Goodness of Fit of the selected PDF models to the date and choose the most appropriate model that represents the dataset.

Several PDF models that are in common use in probabilistic and statistical research are described in this section. These models have mathematical expressions which describes the distribution of random variable. These PDF models are referred to common PDF models throughout the thesis.

Gaussian Distribution

The Gaussian distribution is perhaps the most widely used probability distribution, given by

$$p_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \tag{1-8}$$

where the distribution parameters are μ and σ, the mean and standard deviation of the data, respectively, of the variable. The skewness of Gaussian distribution is 0, which means Gaussian distribution is symmetric about the mean value. The kurtosis of Gaussian distribution is 3.

Lognormal Distribution

A random variable X has a lognormal distribution if $\ln X$ (the natural logarithm of X) is normally distributed. In this case, the probability density function of X is given by

$$p_X(x) = \frac{1}{\sqrt{2\pi}\zeta \cdot x} \exp\left[-\frac{1}{2}\left(\frac{\ln x - \lambda}{\zeta}\right)^2\right] \tag{1-9}$$

where $\lambda \equiv E(\ln X)$ and $\zeta \equiv \sqrt{Var(\ln X)}$, are the mean and standard deviation of $\ln X$, respectively.

For distributions with negative values and positive skewness, a common practice is to shift the data to make sure that all the shifted data is positive. The shifted lognormal distribution is

$$p_X(x) = \frac{1}{\sqrt{2\pi}\zeta \cdot (x-\mu)} \exp\left[-\frac{1}{2}\left(\frac{\ln(x-\mu) - \lambda}{\zeta}\right)^2\right] \tag{1-10}$$

where μ is location parameter, or the amount by which the data must be shifted..

For a dataset which has negative values and negative skewness, it is necessary to first mirror the dataset and then shift the data for better modeling the PDF of the dataset when using lognormal distribution.

Gamma Distribution

The PDF of gamma distribution is given by

$$p_X(x) = \frac{\left(\frac{x}{\beta}\right)^{\gamma-1}}{\beta \cdot \Gamma(\gamma)} \exp\left(-\frac{x}{\beta}\right) \tag{1-11}$$

where β and γ are scale and shape factor, respectively, $\Gamma(\ )$is the gamma function, which is given by $\Gamma(\gamma) = \int_0^\infty t^{\gamma-1} e^{-t} dt$.

Since the variable of gamma distribution must be positive, the same technique used for lognormal distribution must be applied to a dataset which has negative values or negative skewness. The shifted gamma distribution is

$$p_X(x) = \frac{\left(\frac{x-\mu}{\beta}\right)^{\gamma-1}}{\beta \cdot \Gamma(\gamma)} \exp\left(-\frac{x-\mu}{\beta}\right) \tag{1-12}$$

where μ is the location parameter.

Gamma distribution has been identify as an appropriate PDFs for non-Gaussian wind load and this PDF model has been used to predict extreme value of wind loads (Sadek & Simiu(2002).

General Extreme Value Distribution

In extreme value theory, the generalized extreme value distribution (GEV) is a family of continuous probability distributions that are developed within extreme value theory. The expression for the generalized extreme value distribution is:

$$p_X(x) = \frac{1}{\sigma}\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{(-1/\xi)-1} \exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right\} \tag{1-13}$$

for $1+\xi(x-\mu)/\sigma > 0$, where μ is the location parameter, $\sigma > 0$ is the scale parameter and $\xi \in R$ is the shape parameter. The shape parameter ξ governs the tail behavior of

the distribution. The sub-families defined by $\xi \to 0$, $\xi > 0$ and $\xi < 0$ correspond to the Gumbel, Fréchet and Reverse Weibull distribution, respectively.

The three cases covered by the generalized extreme value distribution are often referred to as the Types I, II, and III, which are sometimes also referred to as the Gumbel, Frechet, and Reverse Weibull types, though this terminology can be slightly confusing. Each type corresponds to the limiting distribution of the maxima from underlying distributions. Parent distributions whose tails decrease exponentially, such as the normal distribution, leads to the Type I GEV. Parent distributions whose tails decrease as a polynomial, such as Student's *t*, lead to the Type II GEV. Parent distributions whose tails are finite, such as the beta, lead to the Type III GEV.

Characteristics of Wind Pressure

Wind Pressure Coefficients

The basic definition of a pressure coefficient for a bluff body is given bellow (Holmes 2001):

$$C_p(t) = \frac{p(t) - p_0}{\frac{1}{2}\rho_a \overline{U}^2} \tag{2-1}$$

Where $p(t)$ is the pressure time history at a particular location, p_0 is static reference pressure, ρ_a is density of air, and $\overline{U}$ is the time-averaged velocity measured at an appropriate defined reference height. In the atmospheric boundary layer, the mean wind speed with height, and for the case of low-rise buildings, the reference height for measuring wind velocity is usually taken as either the mean roof height, height to eave or height to roof ridge. The static reference pressure must be taken away from the direct influence of the building.

Four characteristic values of the pressure coefficient are frequently used to establish variability. They are: $\overline{C}_p$, the mean or time-averaged pressure coefficient; C'_p, the root-mean-squared (RMS) fluctuating value, or standard deviation, representing the averaged deviation from the mean; $\hat{C}_p$, the maximum value of the pressure coefficient in a given time period; $\breve{C}_p$, the minimum value of the pressure coefficient in a given time period (Holmes 2001).

Wind Flow over Low-Rise Buildings

Wind flow around a bluff body is complex because of the stochastic features of wind field and the nonlinear interactions between the upstream turbulence and mean velocity gradients with body-generated turbulence. No general analytical theory exists to predict the variation of wind pressure on low-rise buildings. Wind pressure acting on a building depends highly on flow pattern around the building, which in turn is influenced by building configuration, upstream terrain, wind direction and surrounding environment.

The wind pressure fluctuations on a building with time can be attributed to two sources: pressure fluctuation caused by the fluctuations of upwind turbulent velocity and pressure fluctuation produced by vortex shedding, and other unsteady flow phenomena.

Winds parallel or normal to the ridge line of a gable roof building

Shown in Figure 2-5, wind flows perpendicularly to the ridge line over a building with a low-pitched roof. The flow pattern is characterized by separation and reattachment zones, and the formation of separation zones or “bubbles”, depending on the dimensions of the building. The deflection and acceleration of the wind cause the flow to separate at the top of the windward wall. The separation zone is bounded by a free shear layer, a region of high velocity gradients, and high turbulence. The layer is

unstable and will roll up towards the wake, to form concentrated vortices, which are subsequently shed downwind. The effect of turbulence in the approaching flow is to cause the vortices to roll up closer to the leading edge, and a shorter distance to the reattachment(Holmes 2001).

Based on full scale and wind tunnel studies, when the mean wind direction is parallel to the ridge line, the roof could also be seen as aerodynamically flat. The largest wind loads on low rise buildings are usually experienced on the roof close to windward edge in regions of flow separation. Lee and Ho (1990) described the flow mechanisms in the regions of flow separation. When the freestream flow is perpendicular to the line of the separation, a 2D separation bubble forms. The separated shear layer rolls up into discrete vortices and the vortices is then convected by the freestream flow as a whole. Ginger and Letchford (1993) found that the flow mechanisms that generate these pressures are the 2D separation bubble for flow perpendicular to the edge discontinuity.

Cornering wind

For cornering wind, the flow pattern is more complex as shown in Figure 2-6. When the wind blows obliquely to the corner of a roof and has a significant component of the free stream velocity in the direction of the edge discontinuity, 3D conical vortices are formed close to the edge. The highest negative peak pressures locate underneath the vortices.

Since the flow patterns of wind change with respect to different wind angles, the probabilistic characteristics of wind pressures under different wind angles is worthy of study to give us insight to relationship between the flow pattern and the variations of statistical properties of wind pressures.

Non-Gaussian Characteristics of Wind Pressure

Davenport (1961) introduced statistical concepts into wind engineering, assuming that the distributions of wind speed, and wind pressure coefficients are Gaussian distributed. This assumption is based on central limit theorem and the fact a large variety of natural phenomena is Gaussian. This assumption is appropriate when calculating the aggregate effect of wind acting on a large area. However, in the flow separation zones, strongly non-Gaussian effects are correlated and the central limit theorem does not apply. Thus the roof components experience much larger wind pressures than are predicted using an assumption of Gaussian distribution.

Considerable research has been done to better understand the probabilistic characteristics of wind pressure. Kareem (1978) found that the PDFs of wind pressure are skewed in the regions where the mean wind pressure coefficient is less than -0.25, and as stated earlier, this non-Gaussian effect results in a higher rate of occurrence of large fluctuating wind pressures than would be predicted by Gaussian distribution. In other words, the probabilistic distribution of wind pressure deviates from Gaussian distribution in the tail region. Holmes (1981) found that non-linear relationships between surface pressure and wind velocity result in pressure probability distributions which depart significantly from Gaussian. Other studies have also addressed the influence of non-Gaussian nature of local pressure fluctuation.

Common PDF Models for Wind Pressure

Stathopoulos(1980) determined the probability distributions of wind pressures on small scale model of low-rise building and he reported that a Weibull model fits the data adequately. Assuming the wind velocity has a Gaussian PDF, Holmes(1981) established a PDF model of pressure fluctuations, however, this PDF model is only

suitable for weak non-Gaussian situations. Reed (1993) conducted numerical simulations of non-Gaussian and Gaussian pressure fluctuations for glass cladding in tall buildings. The study concluded that significantly greater glass cladding damage would occur using the non-Gaussian assumption. Li et al(1999) compared Gumbel, Weibull and Lognormal PDFs of pressure fluctuations and found that the shifted Lognormal distribution best represents the PDFs of fluctuating pressures. Using probability plot correlation coefficient (PPCC) method, Sadek & Simiu(2002) identified the 3-paramater Gamma distribution as the most appropriate distribution of internal force form among a set of candidate families of distributions (Exponential, Normal, Lognormal, Gumbel, Weibull and Gamma distributions). Holmes and Cochran(2003) found that generalized extreme distribution (GEV) with a positive shape factor fits the extreme value of wind pressure coefficients well.

These studies and others described and used the common PDF models to fit experimental pressure data. However, these PDF models can have difficulty in modeling the distribution of data that is strongly non-Gaussian. Additionally, since the probabilistic content of wind pressure varies dramatically from region to region, and with wind azimuth, roof pitch etc, different PDF models are needed to model the wind pressures in the different locations or conditions, e.g., Cope et al. (2005). This practice is highly inefficient, particularly for strongly non-Gaussian data, like wind pressures on a building surface.

PDF Models Based on Higher-Order Moments

The approach for approximating the PDF of a random variable using limited moments is well known in statistics. The generic problem of determining the shape parameters of a PDF from its limited moments is called the “Method of Moments”.

During the past decades, several moment-based methods have been developed and refined to deal with this class of problem. These methods include Pade-Laplace Method, Hermite Polynomial Method and Maximum Entropy Method, which are the main focus of the thesis. The main work of the thesis is to further develop these methods, explore applicability of these models for non-Gaussian data and conduct a comparison of the accuracy of these models and common PDF models.

The Pade-Laplace Method (PLM) is based on the use of Pade approximation and the Laplace transformation. This method has been applied in different contexts, such as chemistry (Aubard et al. 1987), Molecular Biology (Bajzer et al. 1989), Electronic Engineering (Jay, Ovarlez and Duvaut 2000) and Economics (Kumar 1992), etc. Amidavar and Ritcey (1994) first reported the idea of approximating PDFs using Pade approximation. One limitation of this method is that PLM can only model the PDF of a positive random variable, because the Laplace transformation can only be performed from zero to positive infinity. In their paper, PLM was applied to approximating some common PDFs, but the application of PLM to modeling the PDFs of experimental data were not studied.

The Maximum Entropy Method (MEM) has long been found to be an effective and flexible tool for approximating PDFs, based on limited moments. Typically, solving the problem demands maximizing the Shannon entropy

$$H(p) = -\int p(x) \ln p(x) dx \tag{2-2}$$

under a set of constraints of moment equations (Agmon, Alhassid and Levine 1979). The MEM approach is an effective and flexible approach for PDF approximation, and includes a family of generalized exponential distributions, such as exponential

distribution, normal, lognormal, gamma, beta and Pareto distribution as special cases(Wu 2003). Establishing the MEM PDF are not achieved for free, since establishing the entropy density involves solving a set of coupled nonlinear equations, which are in the form of infinite integration. This is perhaps one reason that the application of MEM appears to be scarce. Finding an efficient and stable algorithm is of great significance for the application of MEM.

The Hermite Method is a class of PDF modeling methods that is based on the orthogonality of the Hermite series (Winterstein 1988). An early trial of Hermite Method is Gram-Charlier series probability distribution. The distribution is given as:

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma_x} \exp\left[-\frac{(x-\mu_x)^2}{2\sigma_x^{\ 2}}\right] \cdot \left[1 + \sum_1^n h_n H_n\left(\frac{x-\mu_x}{\sigma_x}\right)\right] \tag{2-3}$$

where μ_x and σ_x are mean value and standard deviation of the random variable of interest, respectively. h_n is related to the moments of the random variable, $h_n = \frac{1}{n!} E[H_n(z)]$, where $H_n(z)$ is Hermite series. For example: $H_0(z) = 1$; $H_1(z) = z$; $H_2(z) = z^2 - 1$; $H_3(z) = z^3 - 3z$; and $H_4(z) = z^4 - 6z^2 + 3$. From Eq(2-2) , we can see that the PDF is expressed as the product of a standard normal distribution and Hermite polynomials which are orthogonal to the normal distribution. The coefficients are evaluated from the moments of the random variable.

Since the Hermite distribution based on Gram-Charlier series is given in series form, the PDF expressed in Eq(2-2) usually has negative values at some part of the distribution, especially in the tail regions. Another drawback of this method is its lack of

flexibility. The undesired properties are caused by the mathematical expression of the PDF and the limited numbers of terms in computing the density function.

To overcome negative behavior and to make the method more flexible, Winterstein (1988) used first four terms of Hermite series to functionally transform the Gaussian process to a non-Gaussian process. Using this transformation, both hardening ($\gamma_4 < 3$) and softening ($\gamma_4 > 3$) models were developed and shown to be more flexible than the conventional Gram-Charlier series probability distribution. However, the approximated relationship between the shape parameters h_n and the moments results in impaired accuracy when the values of skewness and kurtosis are large. One of the major contributions of this thesis is to improve the approximate relationship between h_n and the moments, and define the effective region of HPM (admissible PDFs), allowing further application to modeling experimental wind pressure data.

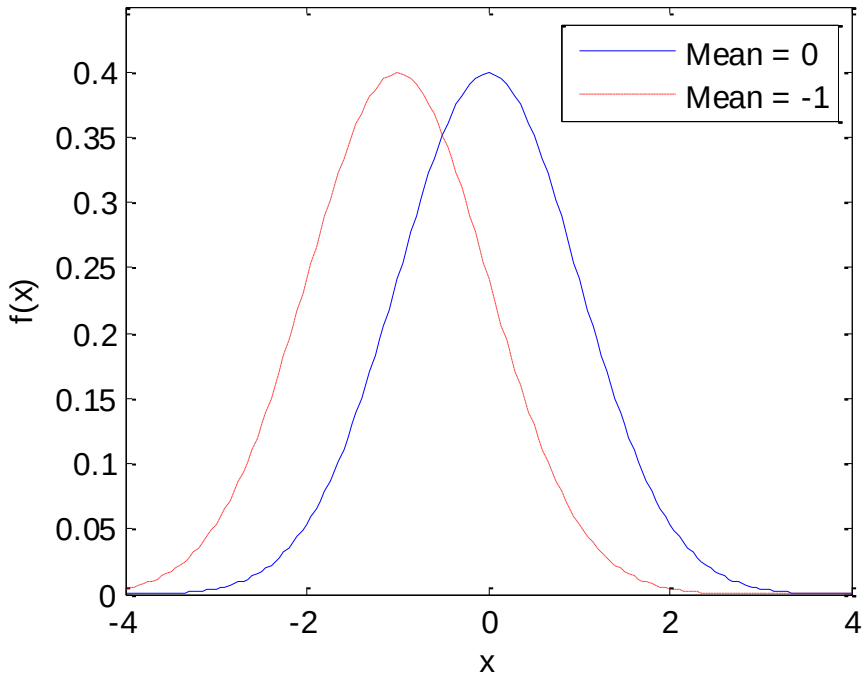


Figure 2-1. Probability density functions with different mean values.

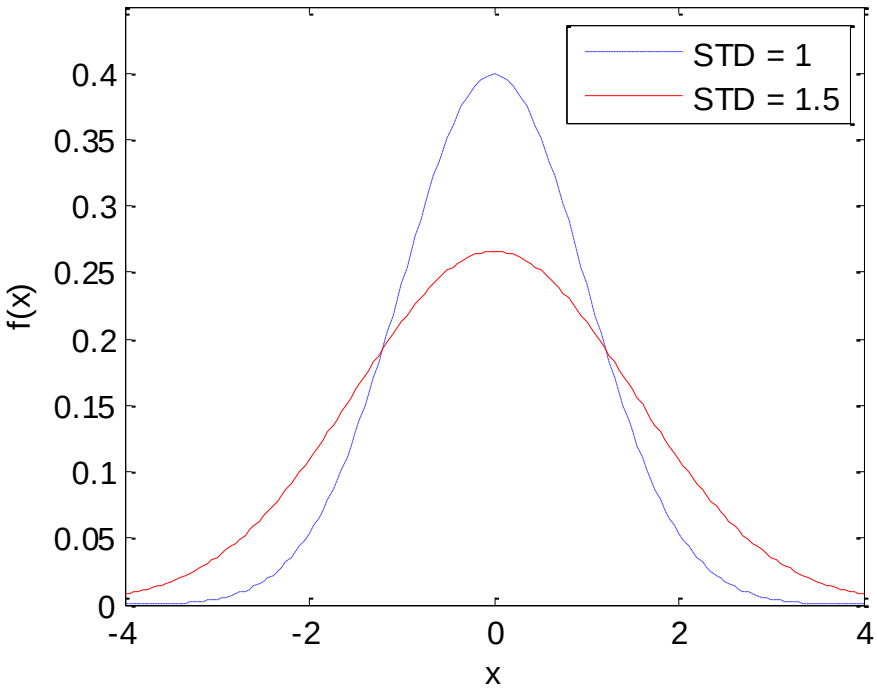


Figure 2-2. Probability density functions with different STD values.

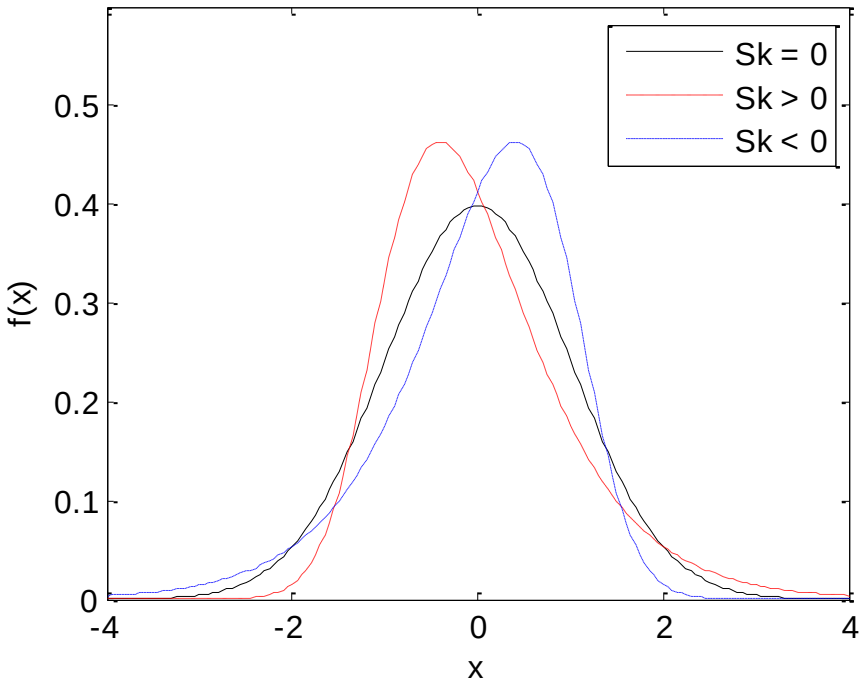


Figure 2-3. Probability density functions with different skewness.

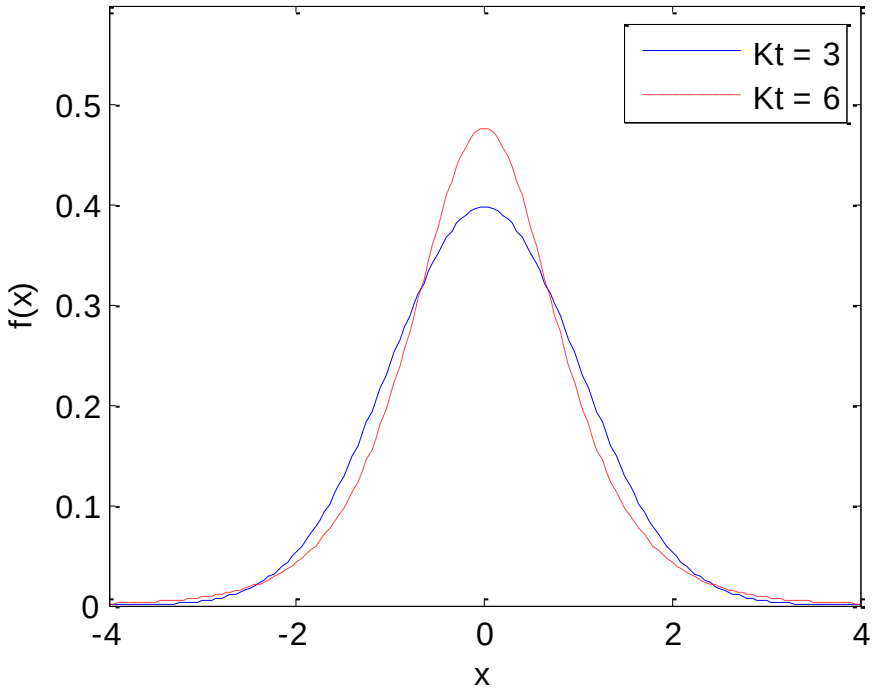


Figure 2-4. Probability density functions with different kurtosis.

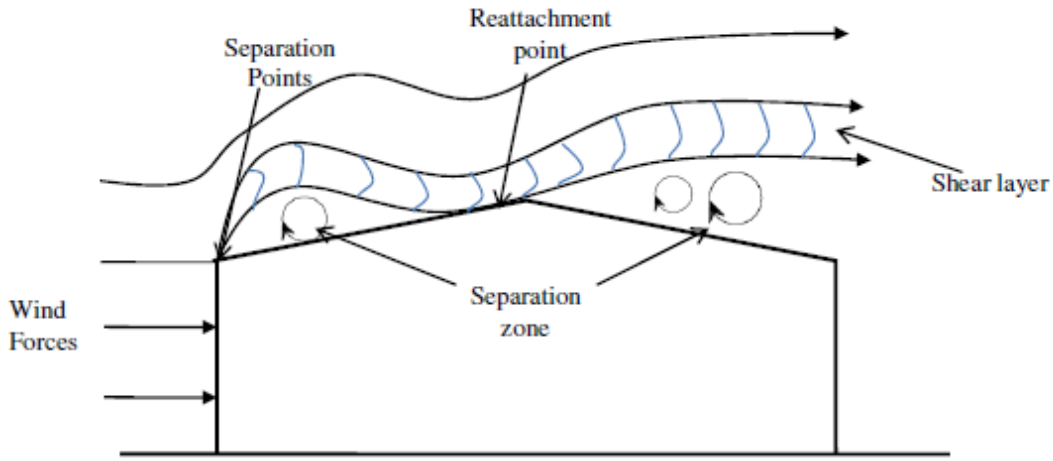


Figure 2-5. Separation and reattachment pattern of wind flow over a low-rise building (Simiu and Miyata 2006)

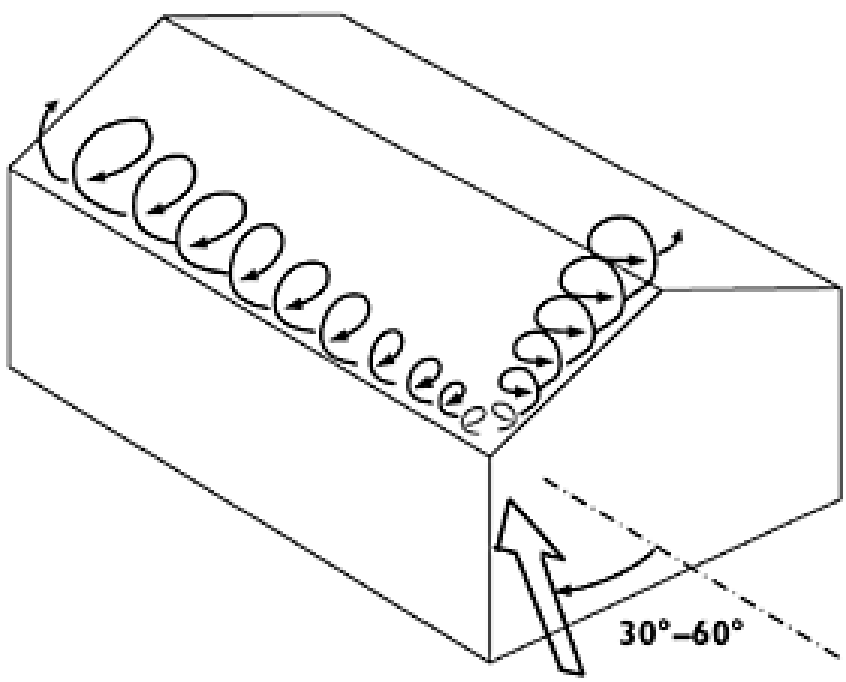


Figure 2-6. Conical vortices for oblique wind directions (Holmes 2001).

CHAPTER 3
WIND PRESSURE DATA AND STATISTICAL PROPERTIES

An extensive database of time histories of fluctuating pressure coefficients were acquired from both scaled wind tunnel and full scale experiments on residential housing shapes. The worst case (peak) loads on these bluff bodies are of interest for the evaluating of design loads. Both the parent distribution (PDF of the full time history) and the distribution of the peak values are non-Gaussian. The severity of the deviation from Gaussian distribution is dependent upon a number of control variables such as wind direction, upwind terrain, and the particular location of interest on the building surface. The accurate modeling of the probability content of these data is the primary goal of this thesis. This chapter describes the wind pressure datasets.

Wind Tunnel Tests

Wind Tunnel Configuration

Wind tunnel measurements of wind pressure on low-rise buildings were conducted in the atmospheric boundary layer wind tunnel at Clemson University's Wind Load Test Facility (WLTF). Atmospheric boundary layer is the lowest part of the atmosphere and its behavior is directly influenced by its contact with earth surface. Trip plates and spires are set up at the entrance to the test section to generate the turbulence in the wind flow. Slant blocks and roughness elements are arranged upwind of the test section and around the model to initiate the velocity profile in the wind flow. These setups combine to develop a simulation of atmospheric boundary layer (Figure 3-1).

Test Model Layout

The tests were performed on a 1:50 scale rectangular plan gable roof model, which is called Clemson standard model (CSM) shown in Figure 3-2A. This acrylic

model is stiff enough to ensure that no significant deformation or vibration occurs. As shown in Figure 3-2B, the model is 14.4 in long and 7.2 in wide with a mean roof height of 3.4 in. The slope of the CSM model is 18.4^0 (4 in 12). Totally 387 pressure taps were installed on the gable roof. As shown in Figure 3-2B, Wind pressure measurement taps were place uniformly in the interior region of the roof at a nominal distance of 1 in. Since the edge regions are of main concern, pressure taps were more densely arranged in the four edge regions of the roof at a distance of 0.2 in.

Wind Pressure Measurement System

The wind tunnel tests were conducted 8 times for five directions: 0^0, 45^0, 90^0, 135^0 and 180^0. For each wind direction, tests were repeated eight times. Each time history was sampled at 300Hz for 120 seconds and filtered to 150 Hz using a digital low-pass filter (Mensah et al. 2011).

Test Data Processing

The measured voltage data were translated into time series of wind pressure by calibration. The effect of the tube system on the measured wind pressure was corrected in the frequency domain using a fast Fourier Transformation. Then, the corrected spectrum of wind pressure was converted back to the time domain (Mensah 2010).

Determining Pressure Coefficient

The pressure time history is then converted to a pressure coefficient history via normalization with a wind speed reference Eq(3-1).

$$C_p(t) = \frac{p_{tap}(t)}{\overline{p}_{ref}} = \frac{p_{tap}(t)}{\frac{1}{2}\rho_a \overline{U}^2} \tag{3-1}$$

where: $C_p(t)$ is the pressure coefficient measured at each pressure tap; $p_{tap}(t)$ is the difference between the model surface pressure measured at a given tap and the reference level static pressure (pressure from the weight of the atmosphere) at a given time t; $\overline{p}_{ref}$ is the mean hourly reference dynamic pressure recorded by a pitot tube at the reference height; ρ_a is the density of air; $\overline{U}$ is the mean velocity of air at the (reference) pitot tube during the sample. Therefore, the pressure coefficients are referenced to the mean pressure (wind speed) at the pitot tube.

Full-Scale Data Collection

In addition to the wind tunnel dataset, full scale data from the Florida Coastal Monitoring Program landfalling hurricane data collection project was also utilized to develop probability models. As shown in Figure 3-3, the dataset was collected on the roof of an occupied residential structure during sustained hurricane force winds. Twenty-four pressure transducers were mounted at corner and edge locations on the roof to measure external dynamic pressures. Time histories of pressure coefficients were then created using an approach similar to that of the wind tunnel dataset. For more details, please refer to Liu et al (2009).

The probability modeling efforts in this thesis utilize both the wind tunnel and full scale datasets as the random processes to be modeled. There is no relationship between the wind tunnel and full scale datasets. Together they represent a fairly comprehensive set of non-Gaussian bluff body surface pressure data created from a variety of control variables (wind direction, upwind terrain, locations of interest on the body surface).

Statistical Properties of Wind Pressure

This section presents some statistical properties of wind pressure coefficients from the scale model wind tunnel experiment for different wind attack angles. Because of the symmetry of the gable-roof model, a total of three cases: winds parallel to ridgeline (wind azimuth 0^0 and$180^{0)}$, cornering winds (wind azimuth 45^0 and 135^0) and winds perpendicular to the ridgeline (wind azimuth 90^0), were considered separately. For each case, the spatial variation of the non-Gaussian features of wind pressure is presented. For the three cases, the contour plots will be presented for azimuth 0^0, 45^0 and 90^0. For azimuth 135^0 and 180^0, the contour plots are provided in appendix A.

Winds Parallel to Ridgeline

Figure 3-4 shows the spatial variation of statistical properties of Cp for parallel winds. The statistics analyzed are peak value of Cps (the minimum Cp value observed in each time history), standard deviation, skewness and kurtosis. The windward edge region is immersed in separation bubble formed by the sharp edge of the roof. At the two windward corners, the mean value, peak value, skewness and kurtosis are highest on the roof, indicating that the two windward corners would experience the most severe wind loads for parallel winds. As the wind passes the roof, the mean and peak values gradually decrease, but skewness and kurtosis tend to increase and approach high values on the ridge zone some distance from windward edge. For parallel winds, the flow on the roof is largely two-dimensional on the roof, except the two windward corners. However, the pitch of the roof seems to disturb the two-dimensional flow, especially at the ridge zone of high skewness and kurtosis.

Cornering Wind

It is well known that the high negative spikes in pressure time series correspond to the negative tail region of the PDF and make the distribution negatively skewed and non-Gaussian. On the other hand, large fluctuations (i.e. high standard deviation) do not necessarily imply strong non-Gaussian effects (high skewness and kurtosis values).

For the cornering wind in Figure 3-5, non-Gaussian properties are much stronger than the two other cases. Under the separation zone, the windward corner experiences high peak values, a large standard deviation and strong non-Gaussian effects. According to Ginger and Letchford (1993), the strong non-Gaussian effects are caused by the formation of 3-dimensional conical vortices. Although the location of high skewness and kurtosis values (strong non-Gaussian effects) starts at the windward corner, the highest skewness and kurtosis values are some distance from the windward corner. The spatial variations of skewness and kurtosis are consistent; the highest skewness region and the highest kurtosis region are superimposed. In this region, the PDFs of wind pressure deviate greatly from Gaussian and represent a challenging region for accurately modeling the PDFs.

Due to the pitch of the roof , the pattern of wind flow is disturbed at the section of the ridge line and windward gable end. Consequently, high peak values, standard deviation and strong non-Gaussian effects are observed at this region. From this, we can see that the roof geometry greatly influence the pattern of flow.

Wind Perpendicular to Ridgeline

Figure 3-6, presents the results for winds approaching perpendicular to the ridgeline. Since the flow separates at the windward edge, high negative peak values are observed there. Because of the pitch of the roof, the flow will separate again on the

ridge, causing large peak pressures and stronger non-Gaussian effects to occur on the leeward roof region. The absolute value of the peaks on the windward roof region is much smaller than on leeward roof region. Skewness is essentially negligible and kurtosis is approximately 4 on windward roof region, indicating the windward roof region could be recognized as a nearly Gaussian region for perpendicular wind. However, on leeward roof region, non-Gaussian effects still exit with skewness largely -1 and kurtosis largely 6.

Uncertainty Associated with Modeling PDFs of Non-Gaussian Data

There are typically three main sources of uncertainty in the problem of modeling PDFs ((Draper 1995, Chatfield 1995)

a) Model uncertainty, uncertainty about the structure of the model;

b) Statistical uncertainty, uncertainty about estimates of the model parameters, assuming that we know the structure of the model;

c) Random variation in observed variables even when we know the structure of the model and the values of the model parameters.

Generally, the errors arising from model uncertainty are worse than those arising from other sources (Chatfield 1995). The common PDF models have only two or three shape parameters, which limit their capacity to capture the shape of non-Gaussian data, especially for strongly non-Gaussian data. To reduce the model uncertainty, it is attractive to develop PDF models which utilize higher moments and thus better capture the shapes of non-Gaussian PDFs.

On the other hand, the randomness embedded in measurement would cause the variation of estimated model parameters. This is statistical uncertainty. Since higher moments are more greatly influenced by rare extreme outcomes than are lower

moments, the statistical uncertainty can be sufficiently large to affect the shape of a distribution.

As it has been introduced, the wind tunnel tests were conducted for five directions: 0^0, 45^0, 90^0, 135^0 and 180^0. For each wind direction, tests were repeated eight times. Each time history was sampled at 300Hz for 120. That is each time history has 36000 data points. To conduct the uncertainty analysis of statistical moments, twenty taps (each tap has eight time histories) are selected for the calculation of statistics, such as mean, standard deviation, skewness and kurtosis, respectively. The values of mean and skewness of wind pressure data are negative, but the absolute values of mean and skewness are taken for convenience of comparison and presentation. For each tap, the four statistics are calculated 8 times from the 8 time histories, and then sorted as shown along the x-axis in the four plots, respectively. The variation of the statistics could be analyzed and displayed on the y-axis. The results are presented in Figure 3-6.

In Figure 3-7, for each box, the central mark is the median (central tendency of the four statistics), the edges of the box are the 25th and 75th percentiles, the whiskers extend to the most extreme data points not considered outliers, and outliers are plotted individually; the edges, whiskers and outliers are indices of uncertainty embedded in statistics. By comparing the four plots, two trends are observed. One is that the variations of the four moments increase as the values the four statistics increase, respectively, except that the variation of the mean values first decreases and then increases. Meanwhile, by observing that the scale of vertical axis in the four plots, the uncertainty of skewness and kurtosis is much larger than those of mean and STD; so the other trend is that the higher the statistical moments used, the more uncertain these

moments will be for a given data. The whiskers and outliers of skewness and kurtosis are much more widely deviated from the central tendency than those of mean and STD. It is shown that the statistical uncertainties of skewness and kurtosis increase dramatically (the whiskers and outlier are several times larger than the central tendency of the statistics), when strong non-Gaussian effects ($sk > 2$, $kt > 10$) are encountered. These whiskers and outliers will dramatically influence the value of skewness and kurtosis, which is sufficient to distort the shape of the generated distribution and result in deviation of the generated distribution from the target PDF.

Based on above discussion, it is demonstrated that the variation of skewness and kurtosis (estimated by method of moment) is much larger than the variation of mean and standard deviation. However, a small change in skewness or kurtosis would make the PDFs generated by models deviate dramatically from target distribution, the data histogram. To reduce the uncertainty embedded in skewness and kurtosis, maximum likelihood method is applied to estimate skewness and kurtosis parameters, which maximize the likelihood estimate. The procedure of estimating skewness and kurtosis for Hermite Polynomial Method using maximum likelihood method will be presented in Chapter 6.

Figure 3-1. Wind tunnel test section for 1:50 suburban terrain.

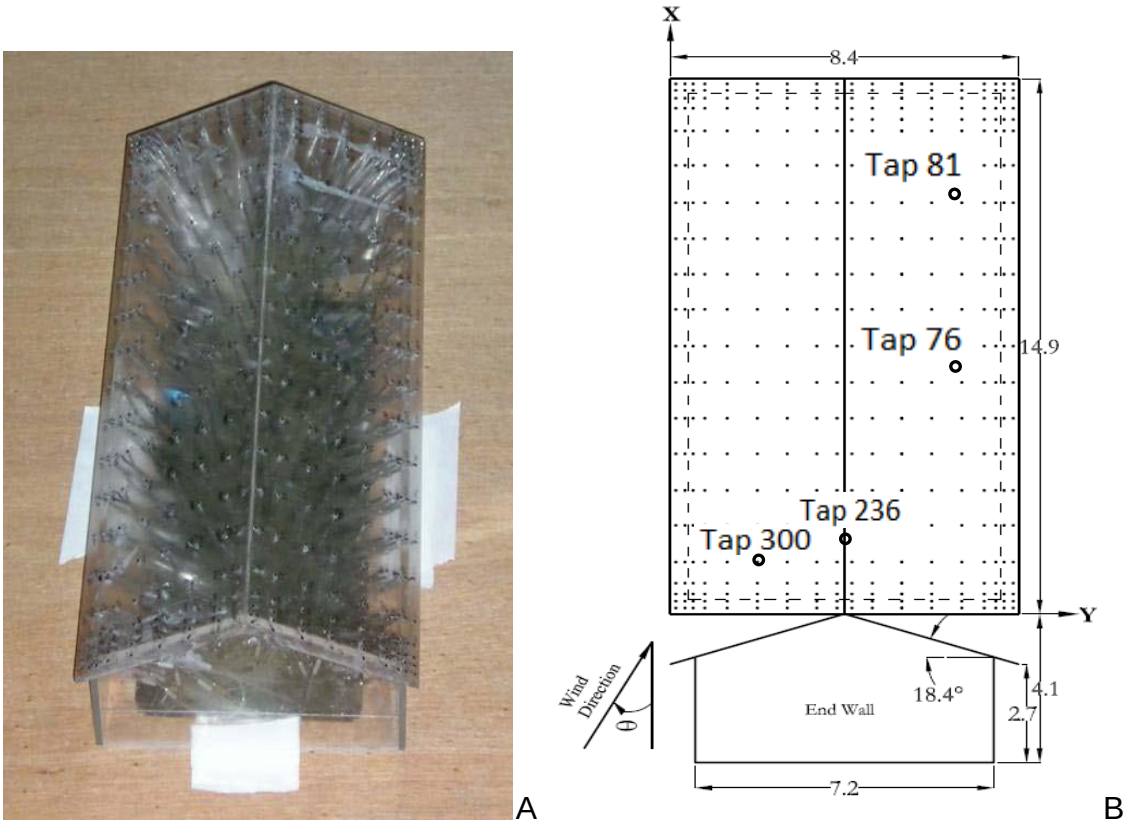


Figure 3-2. Layout of 1:50 Scale house model (CSM 4-12) (Mensah et al. 2011).

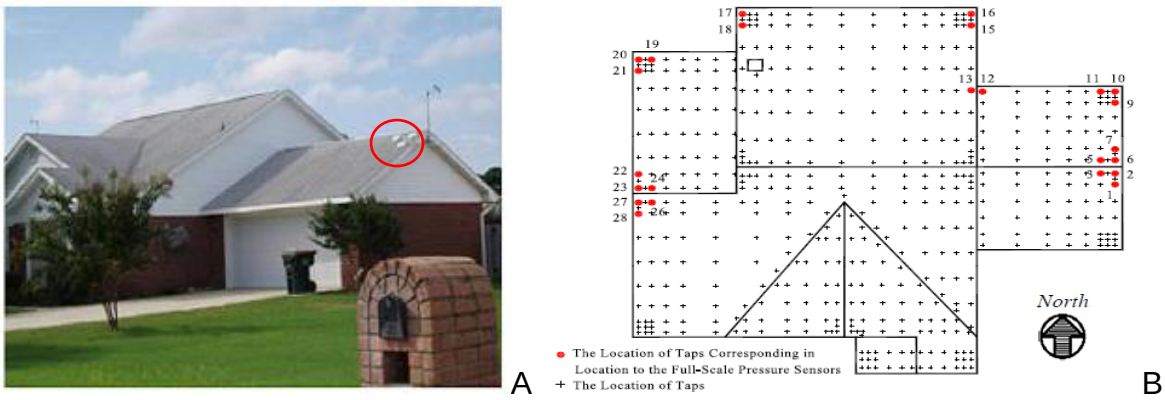


Figure 3-3. Full scale data measurement and tap location: A) anemometer location and pressure sensors, B) roof plan (Liu et al (2009).

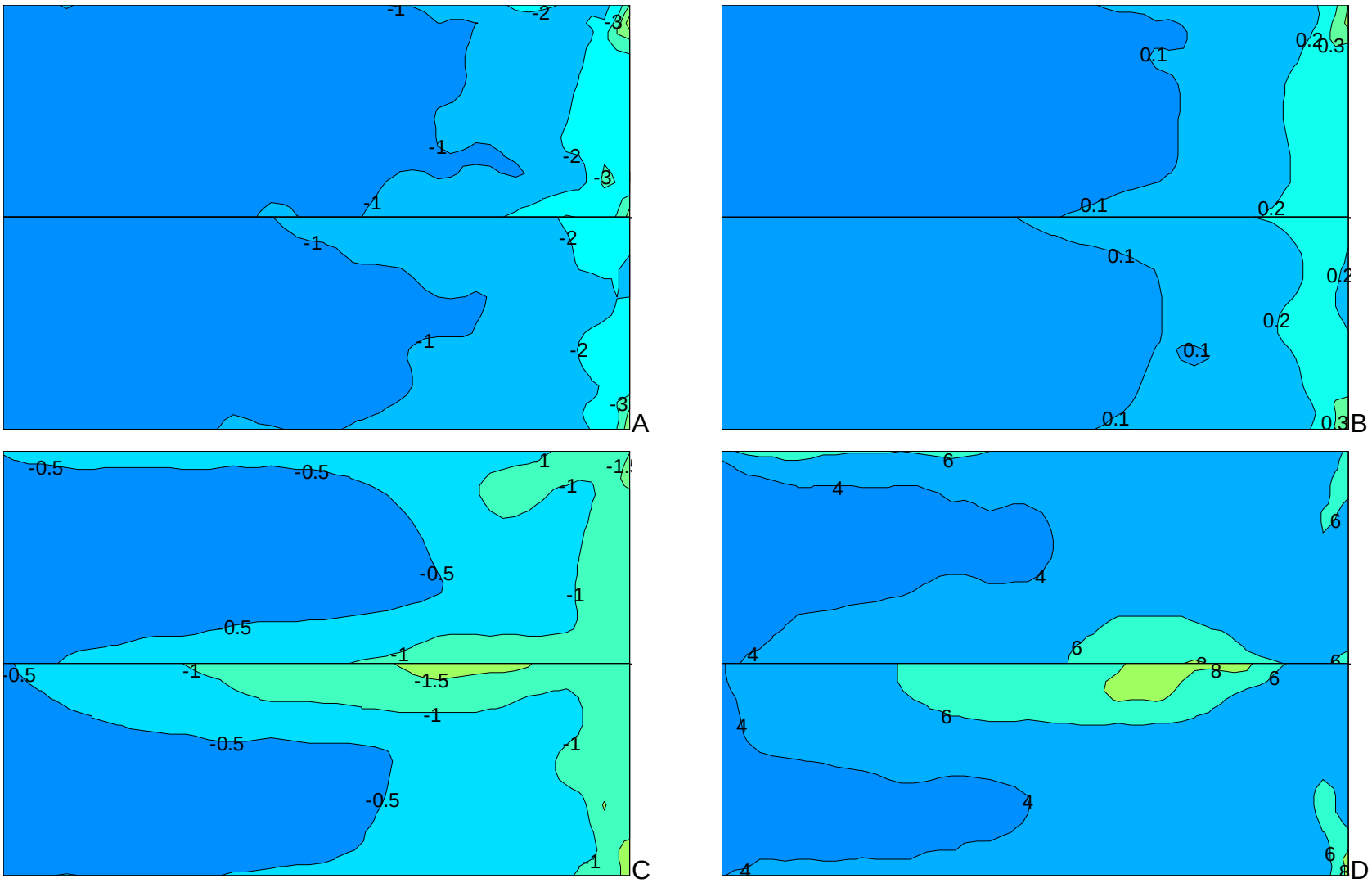


Figure 3-4.Spatial variation of statistical properties of Cps for winds parallel to the ridge. A) peak value, B) standard deviation, C) skewness, D) kurtosis.

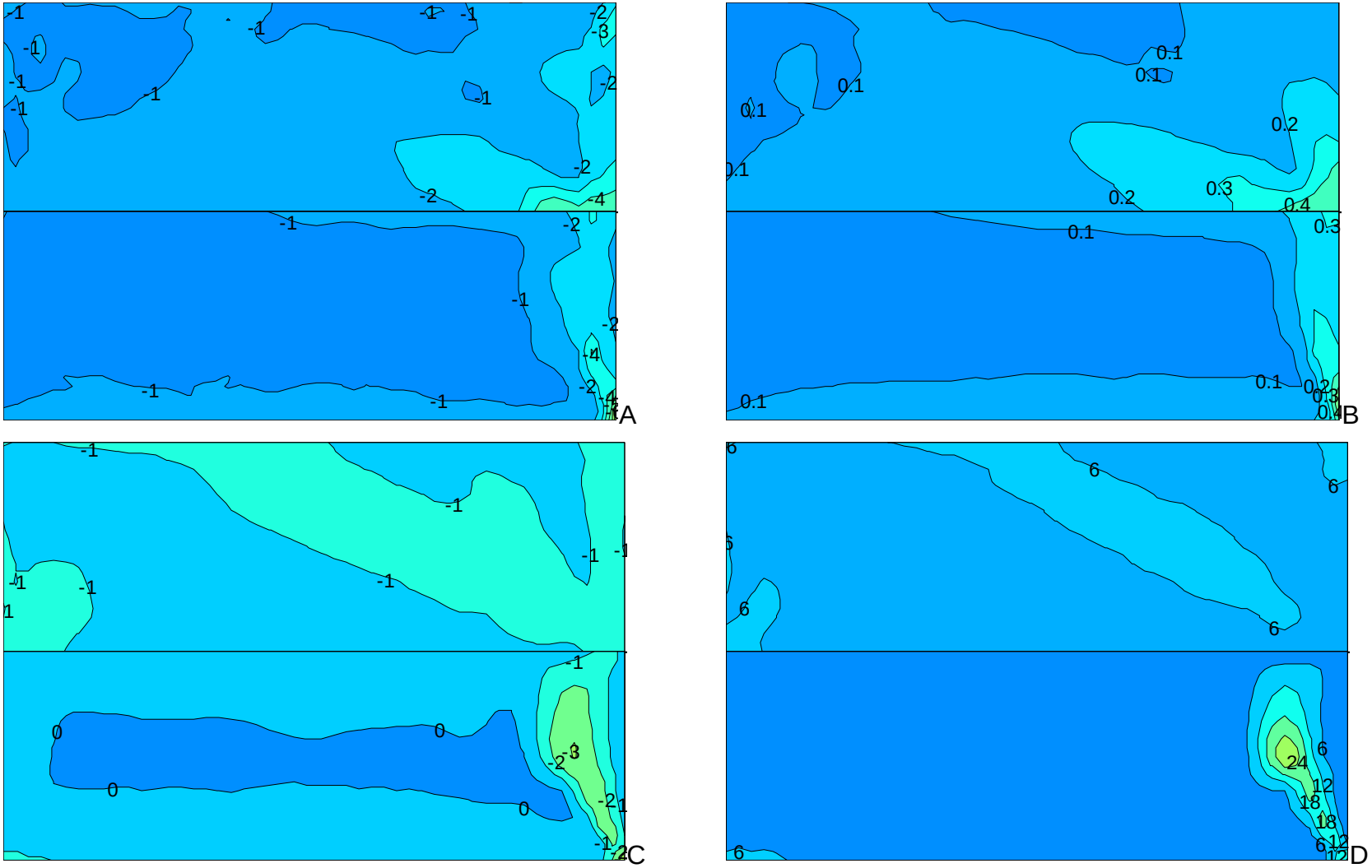


Figure 3-5. Spatial variation of statistical properties of Cps for cornering winds. A) peak value, B) standard deviation, C) skewness, D) kurtosis.

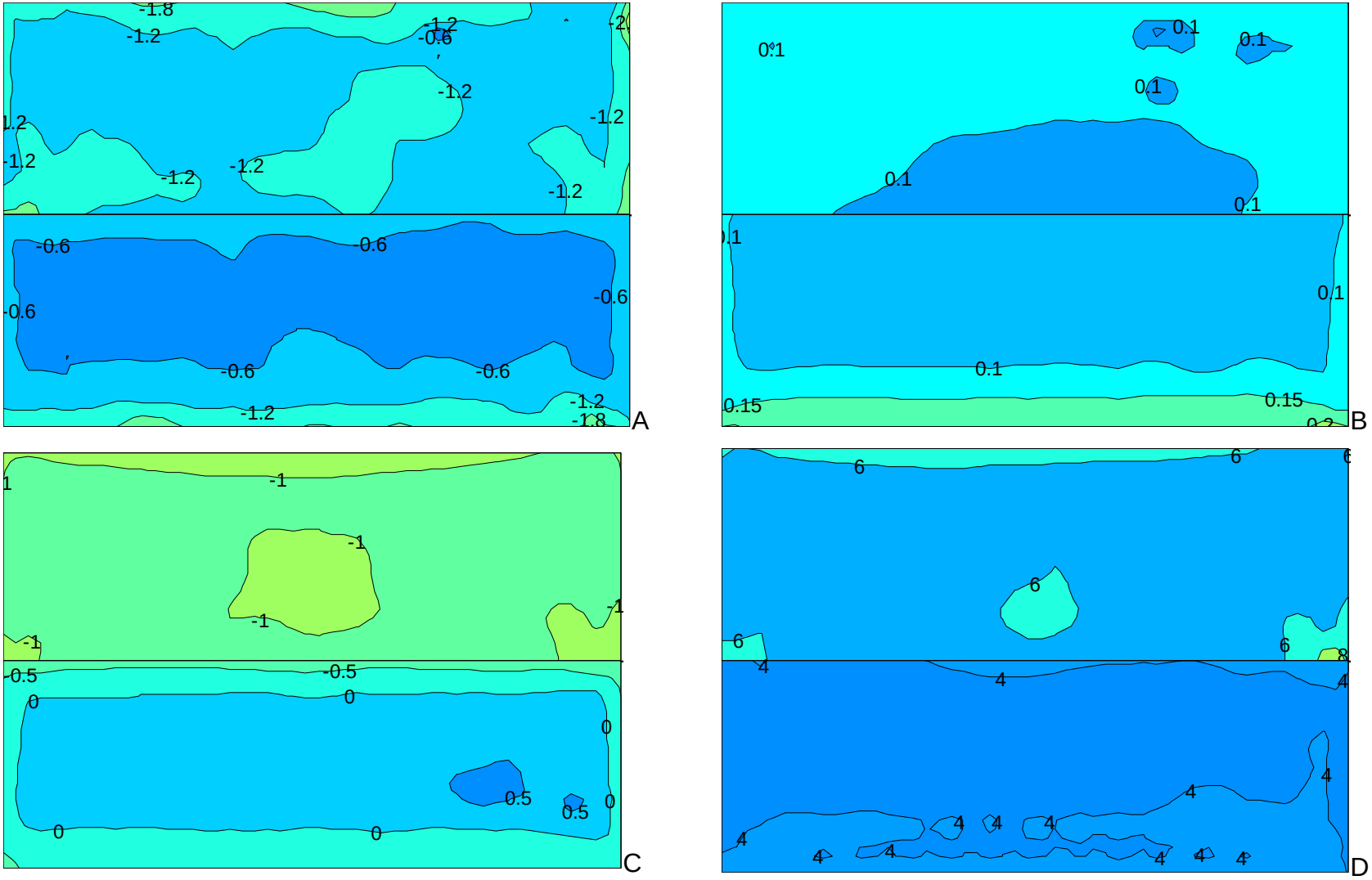


Figure 3-6. Spatial variation of statistical properties of Cps for perpendicular winds. A) peak value, B) standard deviation, C) skewness, D) kurtosis.

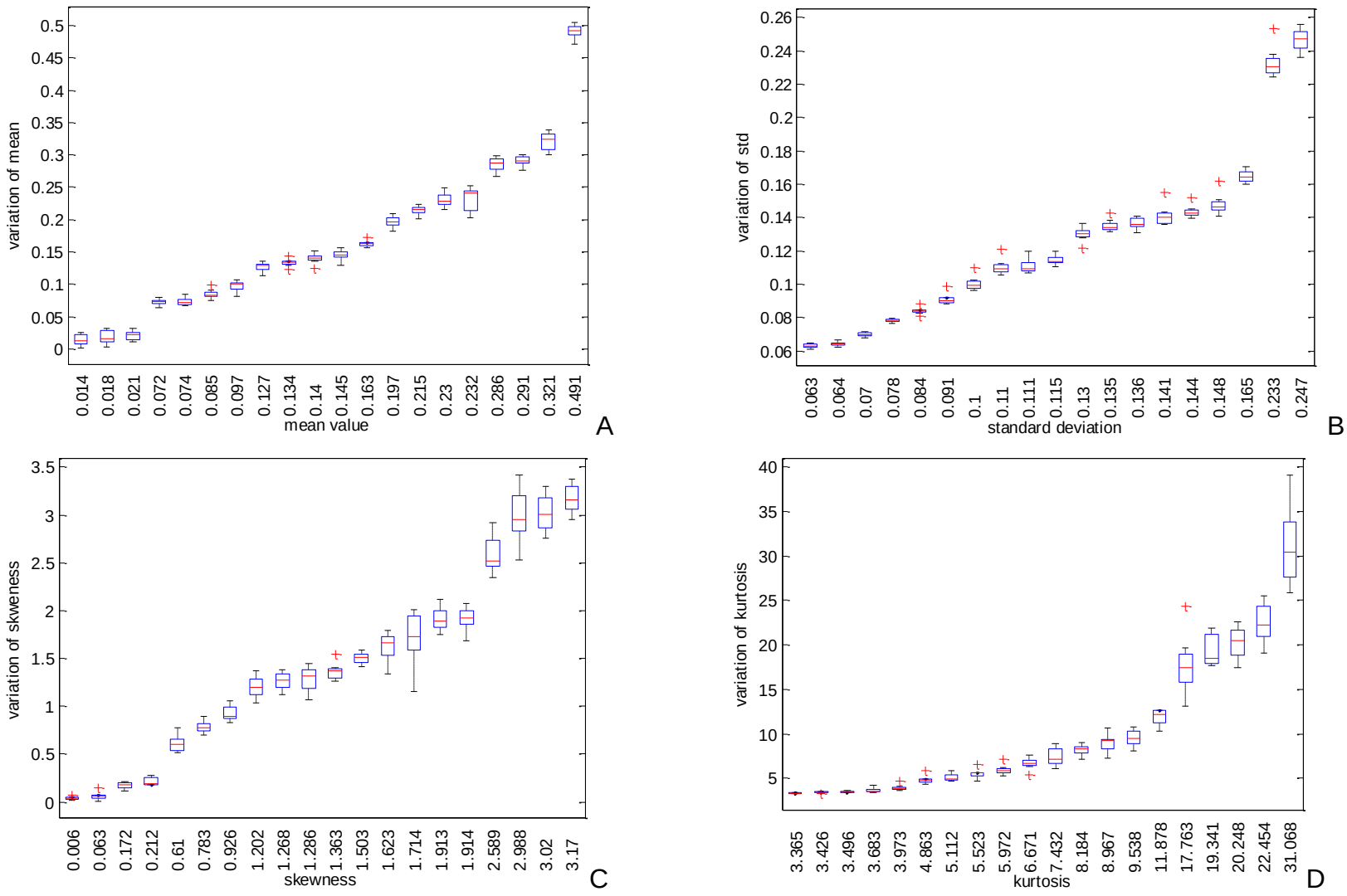


Figure 3-7. Data uncertainty of the first four moments.

CHAPTER 4
PADE LAPLACE METHOD

When a Probability Denstiy Function (PDF) model is selected as an admissible PDF model, the PDF model should satisfy the following conditions:

1) all probability density values calculated from the PDF should be real and non-negative;
2) the integration of the PDF should be unity;

After these methods are formulated or improved, validation of these methods is based on the above two conditions.

Traditional Formulation of Pade Laplace Method

The application of Pade Laplace Method (PLM) to approximating PDFs was first explored by Amidavar and Ritcey (1994). Since the PLM is based on Pade approximation and Laplace transformation, it is necessary to first introduce Pade approximation. The following description of the one-point and two-point Pade approximation are taken from (Amindavar and Ritcey 1994).

One-point Pade Approximation

Let us consider a function $h(u)$. That is analytical about $u = 0$ and could be expanded as Taylor series in one complex variable u :

$$h(u) = \sum_{n=0}^{\infty} c_n u^n , \qquad c_n \in R \tag{4-1}$$

here c_n are the coefficients of Taylor expansion of $h(u)$. This is a formal equality in the sense that the series converges for u within the region of convergence; for large u , the series represents the analytic continuation.

The Pade Approximation (PA) utilizes a rational function to approximate $h(u)$, with denominator degree, M , and numerator degree, L . The PA is constructed in such a way that the coefficients of its expansion in ascending powers of u equals to the coefficients of the expansion of $h(u)$ in ascending powers of u , up to degree $M+L$.

The PA with denominator degree M and numerator degree L , $PA^{[L/M]}(u)$, is defined from the series $h(u)$ as a rational function by

$$PA^{[L/M]}(u) = \frac{\sum_{n=0}^{L} a_n u^n}{\sum_{n=0}^{M} b_n u^n} \tag{4-2}$$

where the coefficients $\{a_n\}$ and $\{b_n\}$ are determined so that

$$\frac{\sum_{n=0}^{L} a_n u^n}{\sum_{n=0}^{M} b_n u^n} = \sum_{n=0}^{L+M} c_n u^n + \mathrm{O}\left(u^{L+M+1}\right) \ , \ u \to 0 \ b_0 = 1 \tag{4-3}$$

where $\mathrm{O}\left(u^{L+M+1}\right)$ takes into account terms of order higher than u^{L+M} .To obtain the coefficients $\{a_n\}$ and $\{b_n\}$, the two sides are equated by matching the coefficients of same powers for $\{1, u, u^2, \cdots\}$. This is the “moments matching approach”,

$$\left\{\sum_{n=0}^{M} b_n u^n\right\}\left\{\sum_{n=0}^{L+M} c_n u^n\right\} = \sum_{n=0}^{L} a_n u^n + \mathrm{O}\left(u^{L+M+1}\right) \tag{4-4}$$

the moment matching conditions can be written explicitly, first for the $\{b_n\}$, then for the $\{a_n\}$,by crossing-multiplying .Taking $b_0 = 1$,without loss of generality, we find that

$$\begin{aligned} &\sum_{n=0}^{M} b_n c_{L-n+j} = 0, && 1 \le j \le M \\ &\sum_{n=0}^{M} b_n c_{L-n+j} = -c_{L+j}, && 1 \le j \le M \end{aligned} \tag{4-5}$$

The equations in (4-5) form a set of M linear equations for the M unknown denominator coefficients. The set is obtained by forming the Hankel matrix and linear system,

$$\begin{bmatrix} c_{L-M+1} & c_{L-M+1} & \cdots & c_{L-M+1} \\ \vdots & \vdots & \vdots & \vdots \\ c_{L-M+1} & c_{L-M+1} & \cdots & c_{L-M+1} \\ \vdots & \vdots & \vdots & \vdots \\ c_{L-M+1} & c_{L-M+1} & \cdots & c_{L-M+1} \end{bmatrix} \cdot \begin{bmatrix} b_M \\ \\ b_k \\ \\ b_1 \end{bmatrix} = - \begin{bmatrix} c_{L+1} \\ \\ c_{L+k+1} \\ \\ c_{L+M} \end{bmatrix} \tag{4-6}$$

which is solved for the denominator coefficients. The numerator coefficients are now determined form Eq(4-4), by back-substitution,

$$a_j = c_j + \sum_{i=1}^{\min(M,j)} b_i c_{j-i} \quad , \qquad 1 \le j \le L \tag{4-7}$$

Now Eq(4-6) and Eq(4-7) determine the PA for a function $h(u)$ expanded in Taylor series.

Two-point Pade Approximation

The PA introduced previously, is a one-point PA, since they match a single power series about only one particular point in the u plane. Now, we consider a PA which simultaneously matches power series expansions at two points. For application to positive random variables, we consider analytic functions that have expansions at $u = 0$ and $u = \infty$ as

$$h(u) = \sum_{n=0}^{\infty} c_n u^n , \qquad u \to 0 \tag{4-8}$$

$$h(u) = \sum_{n=0}^{\infty} d_n u^{-n} , \qquad u \to \infty \tag{4-9}$$

where d_n is coefficients of expansion. It is assumed that $h(u)$ is single valued and that it has no singularities on the positive real axis, save positive at $u=0$ and $u=\infty$.

A two-point PA to $h(u)$ is a rational function, $PA_{[J/K]}^{[L/M]}(u)=A(u)/B(u)$, where $B_0=1$. $A(u)$ and $B(u)$ are polynomials of degree L and M, respectively. The $L+M+1$ coefficients of the approximation are chosen to make the first $J+1$ terms of Eq(4-8) and K terms of Eq(4-9) agree with those of the approximant. The numerator and denominator coefficients are such that when the rational function is expanded as power series in u and u^{-1}, there is agreement with those in Eq(4-8) and Eq(4-9). We formulate the PA so that

$$h(u)-A(u)/B(u)=\mathrm{O}\left(u^{J}\right) \quad , \quad u\to 0 \tag{4-10}$$

$$h(u)-A(u)/B(u)=\mathrm{O}\left(u^{-K}\right) \quad , \quad u\to \infty \tag{4-11}$$

By matching the coefficients of expansions of power series of u and u^{-1}, the coefficients in $A(u)$ and $B(u)$ could be determined from Eq(4-10) and Eq(4-11)

$$\frac{\sum_{n=0}^{M-1} a_n u^n}{1+\sum_{n=1}^{M} b_n u^n}=\sum_{n=0}^{J} c_n u^n \quad , \; u\to 0 \tag{4-12}$$

$$\frac{\sum_{n=0}^{M-1} a_n u^n}{1+\sum_{n=1}^{M} b_n u^n}=\sum_{n=1}^{K} d_n u^{-n} \quad , \; u\to \infty \tag{4-13}$$

The above set of equations could be further reduced to

$$\begin{aligned} & a_j=\sum_{i=0}^{j} b_i c_{j-i}, && 1\le j\le L \\ & \begin{cases} a_j=0, & j>M-1 \\ b_i=0, & i>M \end{cases} \end{aligned} \tag{4-14}$$

$$a_{M-l-1} = \sum_{i=0}^{l} d_{i+1} b_{M-l+i}, \qquad 0 \le l \le K-1$$
$$\begin{cases} a_j = 0, & j > M-1 \\ b_i = 0, & i > M \end{cases} \tag{4-15}$$

Now the PA is established by matching the coefficients of the powers of u and u^{-1}, given its expansions at $u = 0$ and $u = \infty$.

Approximating PDF Using Pade Approximation

According to Probability Theory and Statistics, the moment-generating function (MGF) of any random variable is an alternative definition of its probability distribution. Thus, it provides the basis of an alternative route to analytical results compared with working directly with probability density functions or cumulative distribution functions. The moments are the relatively easily computed statistics of random variable. They are defined for the PDF of a positive random variable X by

$$\mu_n = \int_0^{+\infty} x^n \cdot f(x) dx \qquad n = 1,2,\cdots, \qquad \mu_0 = 1 \tag{4-16}$$

For notational purposes we adopt $f(x)$ for the PDF, $F(x)$ for the Cumulative Density Function (CDF), and $h(u)$ for the MGF of positive random variable.

$$F(x) = \int_0^x f(x) dx \tag{4-17}$$

$$h(u) = E\left[e^{-ux}\right] = \int_0^{\infty} f(x) e^{-ux} dx \tag{4-18}$$

The reason why PLM concentrates on positive random variables as the low limits of all the integrals is zero.

According to Probability Theory, the PDF could be determined from its MGF, if the MGF is expanded into a power series around one or more points. Then the PA to the MGF in terms of the moments

$$\begin{aligned} h(u) &= \int_0^\infty f(x) e^{-ux} dx \\ &= \int_0^\infty f(x) \sum_{n=0}^{\infty} \frac{(-u)^n}{n!} x^n dx \\ &= \sum_{n=0}^{\infty} \frac{(-u)^n}{n!} \int_0^\infty f(x) x^n dx \\ &= \sum_{n=0}^{\infty} \mu_n \frac{(-u)^n}{n!} \end{aligned} \tag{4-19}$$

In practice, moments of all orders are unavailable, so that we assume knowledge of only finite number N moments, truncating the series after that. We consider Eq(4-19) as asymptotic to $h(u)$

$$h(u) \approx \sum_{n=0}^{N} \mu_n \frac{(-u)^n}{n!} \tag{4-20}$$

The PA extrapolates the moments and forms a rational function to the original MGF described by an expansion limited to N terms. As suggested, it is only for approximants with $L \le M$ that the convergence rate and uniqueness can be assured. To determine the PDF from the MGF, we use residue formula to expand the PA,

$$PA^{[L/M]}(u) = \frac{\sum_{n=0}^{L} a_n u^n}{1 + \sum_{n=1}^{M} b_n u^n} = \sum_{i=1}^{M} \frac{\lambda_i}{u - p_i} \tag{4-21}$$

Then apply Laplace inversion on the right-side of Eq(4-21), we get

$$f(x) = \sum_{i=1}^{M} \lambda_i e^{p_i x} \tag{4-22}$$

The CDF is

$$F(x)=1-\sum_{i=1}^{M}\frac{\lambda_i}{p_i}e^{-p_i x}, \qquad p_i>0 \tag{4-23}$$

If we hope to obtain a more uniform approximation, we can use a two-point PA. This requires the MGF be expanded in the vicinity of more than point, often at $u=0$ and $u=\infty$. To form a two point PA, we need to first write

$$\begin{aligned} h(u) &= \int_0^\infty f(x)e^{-ux}dx \\ &= \int_0^\infty \sum_{n=0}^{\infty}\frac{f^{(n)}(0)}{n!}x^n e^{-ux}dx \\ &= \sum_{n=0}^{\infty}\frac{f^{(n)}(0)}{u^{n+1}} \end{aligned} \tag{4-24}$$

Now, using this expansion and that in Eq(4-20), we construct a two point PA. The truncated summation in Eq(4-24) provides some knowledge about the behavior of the MGF of the variable x for large $|u|$, while Eq(4-20) can be used for small $|u|$. In other words, better approximations for $f(x)$ are obtained as $x\to\infty$ using Eq(20) and when we know more terms of Eq(4-24). This clearly motivates the two-point PA. However, for most of PDFs, getting the derivatives of the PDF at the origin, $f^{(n)}(0)$, is not feasible, because the PDF is our target and the expression of the PDF is often unknown.

Examples of Pade Laplace Method

The previous section introduced the traditional formulation of the Pade-Lapace Method (PLM). For better understanding this method it is necessary to introduce some examples of PLM application. These examples given by Amindavar and Ritcey (1994) are presented in this section and will be further discussed as a validation of the PLM in the next sections.

Example One

Now let us consider a function, $h(u)$, which is

$$h(u) = \frac{1}{u^2 + 2u + 1} \tag{4-25}$$

The Taylor expansion of $h(u)$ is

$$h(u) = 1 - 2u + 3u^2 - 4u^3 + 5u^4 + \cdots, \qquad u \to 0 \tag{4-26}$$

The PA of Eq(4-26) could be determined

$$\begin{aligned}
PA^{[0/1]}(u) &= \frac{1}{2u+1} \\
PA^{[1/0]}(u) &= -2u + 1 \\
PA^{[1/1]}(u) &= \frac{-0.5u + 1}{1.5u + 1} \\
PA^{[0/2]}(u) &= \frac{1}{u^2 + 2u + 1} \\
PA^{[2/1]}(u) &= \frac{\frac{1}{3}u^2 - \frac{2}{3}u + 1}{\frac{4}{3}u + 1} \\
PA^{[0/2]}(u) &= \frac{1}{u^2 + 2u + 1}
\end{aligned} \tag{4-27}$$

From Eq(4-27), we can see that $PA^{[1/2]}(u)$ and $PA^{[0/2]}(u)$ are exactly the same as $h(u)$. This is not surprising result, it rather illustrates the fact that rational functions and their power series are linked through PA.

Example Two

Now consider a function, $h(u)$, which is the MGF of positive random variable

$$h(u) = \frac{e^{(-u/1+u)}}{(u+1)^2} \tag{4-28}$$

The Taylor expansion of $h(u)$ is

$$h(u) = 1 - 3u + 6.5u^2 - 12.16u^3 + 20.88u^4 - 33.76u^5 \cdots, \qquad u \to 0 \tag{4-29}$$

The one-point PA could be obtained by Eq(4-6) and Eq(4-7) using moment matching approach

$$\begin{aligned} PA^{[0/1]}(u) &= \frac{1}{3u+1} \\ PA^{[1/2]}(u) &= \frac{-0.067u+1}{2.3u^2+2.9u+1} \\ PA^{[2/3]}(u) &= \frac{3.74u^2+0.73u+1}{2.0u^3+4.71u^2+3.73u+1} \end{aligned} \tag{4-30}$$

Also the two-point PA is obtained by Eq(4-14) and Eq(4-15). First we need to expand $h(u)$ at $u = 0$ and $u = \infty$. Here we choose $L = 1$, $M = 2$, $J = 2$ and $K = 1$ as an example

$$PA_{[2/1]}^{[1/2]}(u) = \frac{a_1 u + a_0}{b_2 u^2 + b_1 u + b_0} = \begin{cases} c_0 + c_1 u + c_2 u^2 & u \to 0 \\ d_1 u^{-1} & u \to \infty \end{cases} \tag{4-31}$$

Matching the coefficients for $u \to 0$ and $u \to \infty$

$$\begin{aligned} a_0 &= c_0 &&, \quad \textit{matching powers of } u^0 \\ a_1 &= c_0 b_1 + c_1 &&, \quad \textit{matching powers of } u^1 \\ 0 &= c_0 b_2 + c_1 b_1 + c_2 &&, \quad \textit{matching powers of } u^2 \\ a_1 &= b_2 d_1 &&, \quad \textit{matching powers of } u^2 \end{aligned}$$

From Eq(4-28), it is observed that $d_1 = 0$. By substituting coefficients in Eq(4-29), we can obtain $PA_{[2/1]}^{[1/2]}(u)$

$$PA_{[2/1]}^{[1/2]}(u) = \frac{1}{2.5u^2 + 3u + 1} \tag{4-32}$$

The target PDF, which has the MGF $h(u)$, is denoted by $f(x)$. The PDFs, determined by inverse Laplace transformation of $PA^{[1/2]}(u)$ and $PA_{[2/1]}^{[1/2]}(u)$, are denoted by $f^{[1/2]}(x)$ and $f_{[2/1]}^{[1/2]}(x)$, respectively.

$$
\begin{aligned}
&f(x) = \sqrt{x}e^{-(1+x)}I_1\left(2\sqrt{x}\right) \\
&f^{[1/2]}(x) = 2.7e^{-0.64x}\cos(0.17x - 1.58) \\
&f_{[2/1]}^{[1/2]}(x) = 2e^{-0.6x}\sin(0.2x)
\end{aligned}
\tag{4-33}
$$

where $I_1(x)$ is the modified Bessel function of order one.

Application of Pade Laplace Method to Experimental Data

In this section, we apply PLM to wind pressure data derived from the wind tunnel dataset described earlier. Wind pressure coefficients were derived from wind tunnel data developed by Datin and Prevatt (2007) and (Mensah et al. 2011). To facilitate the application of PLM, we synthesize the previous sections by presenting the procedure of PLM.

1. Compute the moments from the dataset;
2. Choose Eq (4-20) or Eq (4-24) to get the coefficients of the expansion;
3. Obtain the approximation to the MGF by either one-point PA or two-point PA;
4. Apply the residue inversion formula on PA;
5. Apply Laplace inversion on the PA to generate the approximated PDF.

As an example of application of PLM to Cp data, the Cp data from Tap 001 with wind direction 0^0 is analyzed. The time history of Cp is shown in Figure 4-2. Since PLM is applicable only to positive random variables and the Cp data is usually skewed to the left, we will first mirror the data and then shift the data to the positive axis for the application of PLM.

The moments are calculated, using $\mu_i = \frac{1}{N}\sum_{j=1}^{N} Cp_j^{\ i}$,

$\mu_1 = 0.5306;$ $\mu_2 = 0.3292;$ $\mu_3 = 0.2380;$ $\mu_4 = 0.2027;$

$\mu_5 = 0.2126;$ $\mu_6 = 0.3020;$ $\mu_7 = 0.6292;$ $\mu_8 = 1.8026;$

The expansion of MGF is

$$h(u) = 1 - 0.5306u + 0.1646u^2 - 0.0397u^3 + 0.0084u^4 - 0.0018u^5 \cdots, \qquad u \to 0$$

Using one-point PA with $L = 1$ and $M = 2$, the PA is calculated as

$$PA^{[1/2]}(u) = \frac{-0.1231u + 1}{0.0516u^2 + 0.4075u + 1}$$

The residual expansion of $PA^{[1/2]}(u)$ is

$$PA^{[1/2]}(u) = \frac{-1.1926 - 7.3866i}{u - (-3.9465 + 1.9484i)} + \frac{-1.1926 + 7.3866i}{u - (-3.9465 - 1.9484i)}$$

Finally by applying inverse Laplace transformation on $PA^{[1/2]}(u)$, the PDF is obtained

$$\begin{aligned} f^{[1/2]}(x) &= (-1.1926 - 7.3866i)\cdot \exp[(-3.9465 + 1.9484i)x] \\ &\quad + (-1.1926 + 7.3866i)\cdot \exp[(-3.9465 - 1.9484i)x]\cdot \\ &= 2\exp(-3.9465x)\cdot[(-1.1926)\cos(1.9484x) - (-7.3866)\sin(1.9484x)] \end{aligned}$$

To see how PDFs change when different L and M values are chosen for PA, the PDFs is calculated using a different pairs of L s and M s.

Limitation of Pade Laplace Method

From Figure 4-3 and Figure 4-4, it is shown that the PDFs generated by the Pade Laplace Method have negative values in tail region or at the origin, when PLM is applied to the moments from this experimental dataset. In fact, even when PLM is used to

approximate an analytical PDF like the one in Example 2, negative values appear in the far negative tail region as shown in Figure 4-5.

PLM generates negative values because of the final mathematical expression to approximate the target PDF. To better understand this reason we need to reconsider Eq(4-21) and Eq(4-22)

$$PA^{[L/M]}(u) = \frac{\sum_{n=0}^{L} a_n u^n}{1+\sum_{n=1}^{M} b_n u^n} = \sum_{i=1}^{M} \frac{\lambda_i}{u-p_i} \tag{4-21}$$

$$f(x) = \sum_{i=1}^{M} \lambda_i e^{p_i x} \tag{4-22}$$

In Eq(4-21), p_i is the root of $1+\sum_{n=1}^{M} b_n u^n$ and could be either real number or complex number, λ_i is the residual of $\frac{\sum_{n=0}^{L} a_n u^n}{1+\sum_{n=1}^{M} b_n u^n}$. According to Euler's formula, $1+\sum_{n=1}^{M} b_n u^n = 0$ has M roots and the M roots are symmetric about real axis. So when one of the p_is is complex number, there exists another p_i which is conjugate to the root. The pair of p_i and λ_i could have the form bellow

$$\begin{aligned} p_{1,2} &= a \pm bi \\ \lambda_{1,2} &= c \pm di \end{aligned} \tag{4-34}$$

For simplicity, assume $M = 2$, that is to say $f(x)$ is the sum of two terms. Plug Eq(4-34) into Eq(4-22)

$$
\begin{aligned}
f(x) &= \lambda_1 e^{p_{1i}x} + \lambda_2 e^{p_2 x} \\
&= (c+di)e^{(a+bi)x} + (c-di)e^{(a-bi)x} \\
&= e^{ax}\left[(c+di)e^{bix} + (c-di)e^{-bix}\right] \\
&= e^{ax}\left[2c\frac{e^{bix}+e^{-bix}}{2} - 2d\frac{e^{bix}-e^{-bix}}{2i}\right] \\
&= 2e^{ax}\left[c\cos(bx) - d\sin(bx)\right]
\end{aligned}
\tag{4-35}
$$

From Eq(4-35) we can see that $f(x)$ is the sum of triangular functions, which have negative value periodically. This explains the negative value in the tail region of PDFs generated by PLM. When $M > 2$, the above derivation still applies.

Based on above discussion, the negative value of generated PDF is a severe restriction of PLM to serve as a robust model for non-Gaussian data where the accurate modeling of the tail region is of pimary importance.

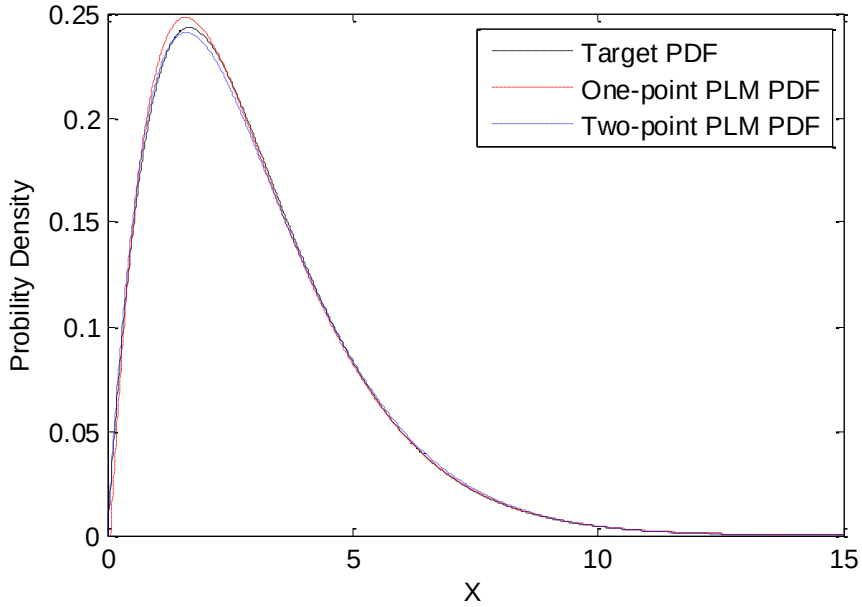


Figure 4-1. Comparison of target PDF with PDFs generated by one-point and two-point PLM.

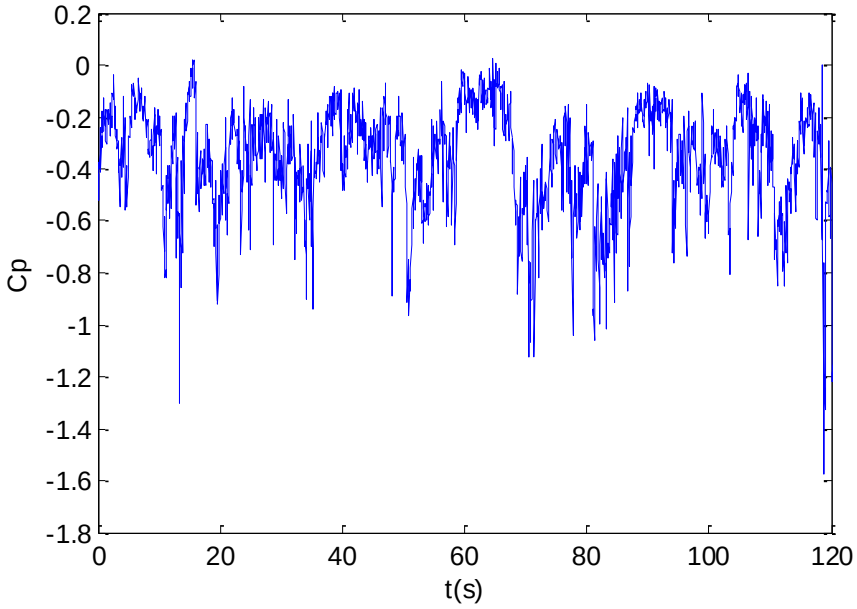


Figure 4-2. Time history of Tap 001.

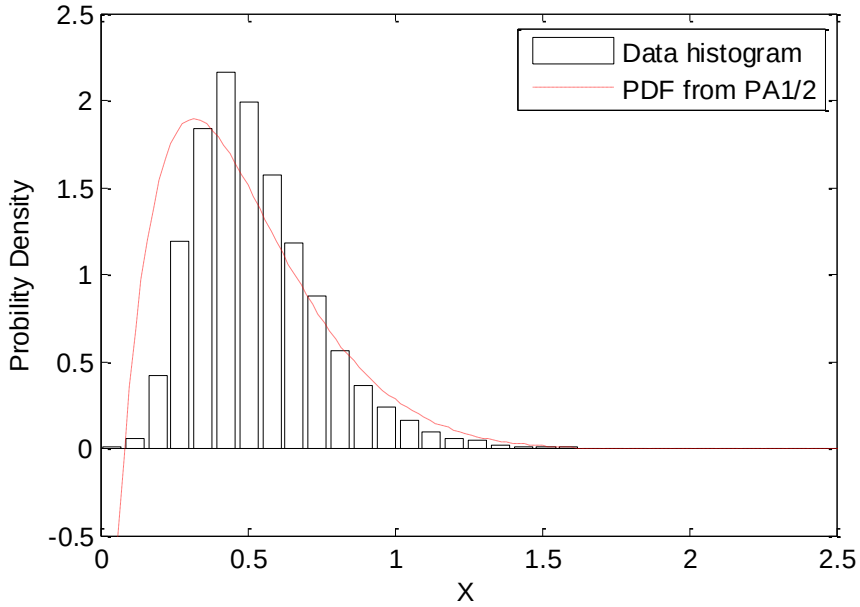


Figure 4-3. Approximating data histogram by $PA^{[1/2]}$

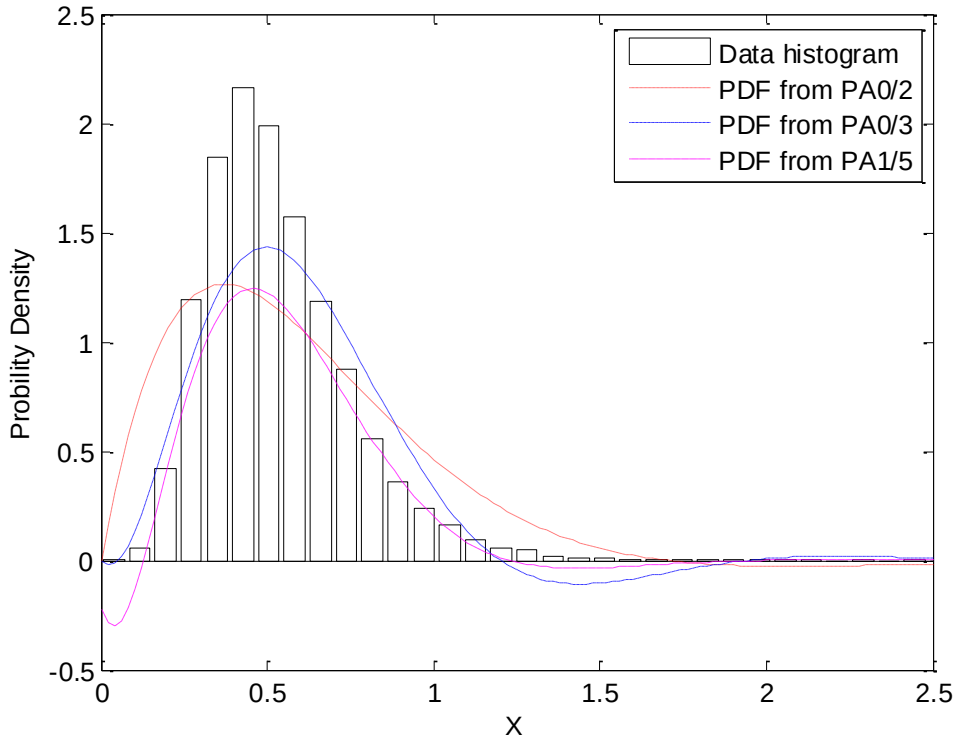


Figure 4-4. Approximating data histogram using different forms of Pade approximation.

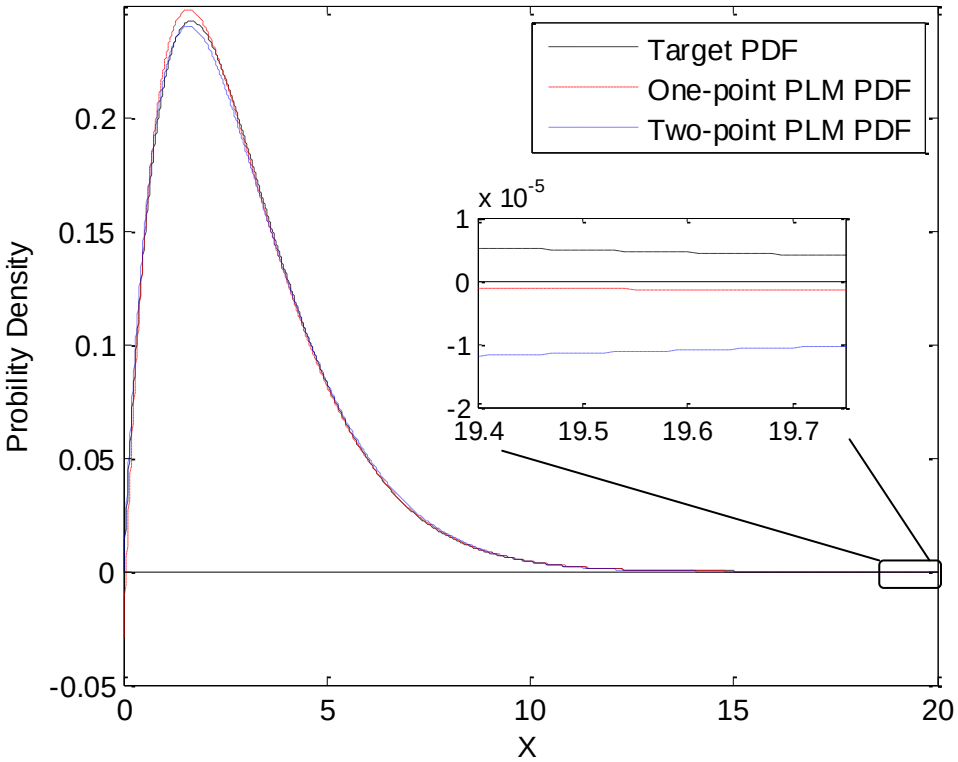


Figure 4-5. Negative values of PDFs generated by PLM in the right tail region.

CHAPTER 5
MAXIMUM ENTROPY METHOD

Traditional Formulation

In the context of physics and information theory, entropy is a measure of uncertainty associated with a random variable and quantifies the informational content of the associated random phenomenon(Sobczyk and Trebicki 1993).

An approach to approximate the PDF of nonlinear systems or data is the Maximum Entropy Method. The principle of maximum entropy states that, of all the probability density distributions $p(x)$ that satisfy the appropriate moment constrains, one should choose the distribution having the largest entropy value, defined as $H(p) = -\int p(x)\ln p(x)dy$. For a finite amount of moment information, there are an infinite number of admissible probability density functions. According to the principle of maximum entropy, the PDF which maximizes the entropy functional is the least biased estimate for the given moment information among the set of infinite PDFs.

Jaynes (1957) first applied entropy functional for determining an unknown PDFand Sobczyk and Trebicki (1993) extended the maximum entropy method to the general class of stochastic nonlinear systems. The following description of the maximum entropy method is taken from Sobczyk and Trebicki (1993). Consider a system of stochastic equations

$$\frac{dX(t)}{dt} = F(X(t), U(t)) \tag{5-1}$$

where $X(t)$ is an unknown response process and $U(t)$ is a given stochastic excitation. It is also assumed that F admits the existence of a stationary solution.

Most often, the system information about process $X(t)$ is given by moments or by the equations for moments. In general, the available information can be expressed as the expected value of polynomial function of $X(t)$, denoted by $G_{r_1 r_2 \cdots r_n}(x)$:

$$E\left[G_{r_1 r_2 \cdots r_n}(x)\right] = \int G_{r_1 r_2 \cdots r_n}(x) p(x) dx = \beta_{r_1 r_2 \cdots r_n} \tag{5-2}$$

where $\beta_{r_1 r_2 \cdots r_n}$ are given numbers, and $r_i = 0,1,2,\cdots M$, where M is the maximum order or correlation moment.

We can also rearrange Eq(5-2) in the form of an ordered sequence equations,

$$E\left[G_k(x_1, x_2, \cdots x_n)\right] = \beta_k, \qquad k = 1,2,\cdots N \tag{5-3}$$

where N is number of moment equations. An additional constrain is the integration of PDF is unit

$$G_0 = \int p(x) dx - 1 = 0 \tag{5-4}$$

where $p(x)$ is the unknown PDF of the nonlinear system.

From the principle of the maximum entropy, one appropriate PDF of the Eq(5-1) is one which maximizes the entropy functional,

$$H(p) = -\int p(x) \ln p(x) dy \tag{5-5}$$

subjected to constraint equations given in Eq(5-2) and Eq(5-3).

With the application of Lagrange multipliers, the entropy functional takes the following form

$$\overline{H}(p) = -\int p(x) \ln p(x) dx - \lambda_0' G_0 - \left[\sum_{r_1 + r_2 + \cdots r_n}^{M} \lambda_{r_1 r_2 \cdots r_n} \left(\int G_{r_1 r_2 \cdots r_n}(x) p(x) dx - \beta_{r_1 r_2 \cdots r_n}(x) \right) \right] \tag{5-6}$$

where $\lambda_{r_1 r_2 \cdots r_n}$ are Lagrange multipliers. The maximum of H can be determined by different Eq(5-6) with respect to p.

$$\frac{\delta \overline{H}(p)}{\delta p} = 0 \tag{5-7}$$

which is

$$-\ln p(x) - 1 - \lambda_0^{'} - \sum_{r_1 + r_2 + \cdots r_n}^{M} \lambda_{r_1 r_2 \cdots r_n} G_{r_1 r_2 \cdots r_n}(x) = 0 \tag{5-8}$$

Then, the PDF could be expressed as

$$p(x) = \exp\left(-1 - \lambda_0^{'} - \sum_{r_1 + r_2 + \cdots r_n}^{M} \lambda_{r_1 r_2 \cdots r_n} G_{r_1 r_2 \cdots r_n}(x) \right) \tag{5-9}$$

The above expression is the general form of maximum entropy distribution satisfying the system of constrains. When only a finite number of moments are given, Eq(5-9) reduces to

$$p(x) = \exp\left(-1 - \lambda_0^{'} - \sum_{r_1 + r_2 + \cdots r_n}^{M} \lambda_{r_1 r_2 \cdots r_n} x_1^{r_1} x_2^{r_2} \cdots x_n^{r_n} \right) \tag{5-10}$$

When MEM is applied to modeling the PDF of random data, the available information is commonly given in the form of the first four moments. The PDF in Eq(5-10) further reduces to

$$p(x) = \exp\left(-\lambda_0 - \sum_{i=1}^{n} \lambda_i x^i \right) \tag{5-11}$$

subjected to

$$\begin{aligned} &\int p(x) dx = 1 \\ &\int y^i p(x) dy = E\left[x^i\right] = b_i \end{aligned} \tag{5-12}$$

where $\lambda_0 = 1 + \lambda_0^{'}$, and b_i is the moment to the ith, which could be measured from the dataset.

When the system information is in the form of measured data, the first four moments could be calculated from the dataset and the non-Gaussian properties would be accounted Eq(5-11).

Limitation of Maximum Entropy Method in the Tail Regions

Although MEM has been of interest as a convenient and flexible tool for approximating PDFs, entropy-based applications are not in common use. One possible reason is that the solution for the λ values is not stable. When applying the MEM, one is faced with the need to solve implicit, nonlinear equations in order to determine the numerical values of the Lagrange multiplier.

To realize the application of MEM, the numerical solution to Eq(5-11) and Eq(5-12) has been studied by researchers. Preliminary research of the existence of the solution and a numerical solution was presented by Alhassid, Agmon and Levine (1978) and Agmon et al (1979). In their algorithm, the Lagrange multipliers are determined by searching a maxima of concave function, instead of determing the numerical solution of a set of implicit nonlinear equations. However, the method developed by Agmon et al (1979) sometime did not converge under a variety of circumstances. Zellner and Highfield(1988) and Ormoneit and White(1999) applied a Newton search algorithm (Atkinson 1989) to the numerical solution of the Lagrange multipliers. Although their approaches improved the convergence, the MEM convergence region was not clearly defined and the optimization of the four non-linear equations is still needed to derive the solution.

With the introduction of the canonical quartic family by Rockinger and Jondeau (2002), the four Lagrange multipliers are reduced to two parameters, α and β, corresponding to skewness and kurtosis, respectively, resulting in improved efficiency arriving at a solution. More importantly, the convergence region of MEM is clearly defined. Figure 5-1 presents the skewness-kurtosis domain for which PDFs exist (the domain is symmetric with respect to the horizontal axis).

The circles represent those points for which the parameters α and β are computed. The red dots represent those points for which the errors between target skewenss and kurtosis values and the computed values are larger than 10^{-5}. It is shown that the computed skewness and kurtosis have discrepancy with the target skewness and kurtosis in the region of skewness close to zero. In this study, we find that the error is due to the fact that the solved Lagrange multipliers become unstable when the skewness approaches zero. This would not prove to be a serious problem, since for most non-Gaussian data, the skewness of data generally lies out of the inaccurate region of the MEM.

However, with the improvement in the convergence region and simplified algorithm, the tail behavior of the PDFs generated by the MEM puts restrictions on its application in modeling highly skewed data. As shown in Figure 5-2, when the kurtosis is fixed, the PDFs generated by the MEM tend to exhibit multimodal behavior as skewness increases, which is inconsistent with the tail behavior of typical non-Gaussian data. The limitation hampers the application of the MEM to strongly non-Gaussian data where the tail region is of significance. To overcome this disadvantage, an alternative formulation of MEM constraints is explored below.

Alternative Constraints for Maximum Entropy Method

When moments are calculated from the dataset, information about the dataset is extracted. However, the dataset contains more information than we have used and the first four moments are just a part of the information. Also, if we could extract different forms of information which may better reflect the actual characteristics of the dataset, it is possible to utilize this information to generate PDFs, which will better represent the distribution of the dataset.

When the first four moments are calculated from data and applied, the constrains in Eq(5-4) are reduced to

$$E[x], \quad E[x^2], \quad E[x^3], \quad E[x^4] \tag{5-12}$$

To eliminate the multimodal behavior of generated PDFs (i.e. the oscillating behavior in tail region), some the traditional constraints should be replaced by inverse moments (the expectation of inverse of the random variable). And then the combination of traditional moments and inverse moments could be used to approximate target distribution. Based on above argument, Eq(5-12) could revised to include other combinations, for example,

$$E[x], \quad E[x^2], \quad E[x^{-1}], \quad E[x^{-2}] \tag{5-13}$$

$$E[x], \quad E[x^2], \quad E[x^{-2}], \quad E[x^{-3}] \tag{5-14}$$

$$E[x], \quad E[x^2], \quad E[x^{-3}], \quad E[x^{-4}] \tag{5-15}$$

$$E[x], \quad E[x^{-1}], \quad E[x^{-2}], \quad E[x^{-3}] \tag{5-16}$$

$$E[x], \quad E[x^{-2}], \quad E[x^{-3}], \quad E[x^{-4}] \tag{5-17}$$

When the dataset represents a continuous random variable ranging from negative to positive values, in which case zero is within the range of dataset, the inverse of zero

is a problem for alternative constrains. To overcome this problem, the dataset should be shifted so that all the data is either all negative or positive.

Next we study the applicability and performance of MEM using traditional and alternative constraints. In this study, PDFs generated by MEM using traditional constraint are called MEM-TC and PDFs obtained from alternative constraints Eq(5-14) through Eq(5-17) are referred to MEM-AC1, MEM-AC2,...and MEM-AC5. The dataset used in this case is wind pressure coefficient data taken form Tap 076 from wind tunnel test (skewness -1.31, kurtosis 5.35), described in Chapter 3. Since the data range from -1.70 to 0.88, the data are shifted by 1.1×0.08 to the left to make sure that all the data have the same sign (here 1.1 is chosen to ensure the admissible PDFs be generated and how to select this value needs further research).

The histogram and PDFs generated by MEM using traditional and alternative constraints are presented in Figure5-3. It is observed that MEM with traditional and alternative constraints could generate reasonably admissible PDFs, except for the MEM-AC1 constraint. Among the admissible PDFs, MEM-AC4 and MEM-AC5 match the histogram well in the mean region. Also, in the left tail region, MEM-AC4 and MEM-AC5 more closely track the histogram than other PDFs. However, in this example, we can observe that MEM PDFs derived from alternative constraints are still not robust, since one of them (MEM-AC1) did not match the histogram at all. If data from another tap is used, it is expected that one or more of the MEM-AC would also fail, generating extremely high values in the mean region. The MEM-TC models generate admissible PDFs for all taps, although the obtained PDFs might exhibit multimodal in the tail region for highly skewed data.

In summary, MEM with traditional constraints could serve as a robust PDF model, so long as the non-Gaussian data is only mildly skewed. The MEM with alternative constraints is not sufficiently robust for practical applications. To overcome the restrictions of PLM and MEM, another model, the Hermite Polynomial Model (HPM) which is flexible and robust for representation of PDFs of non-Gaussian data, will be presented in the next chapter.

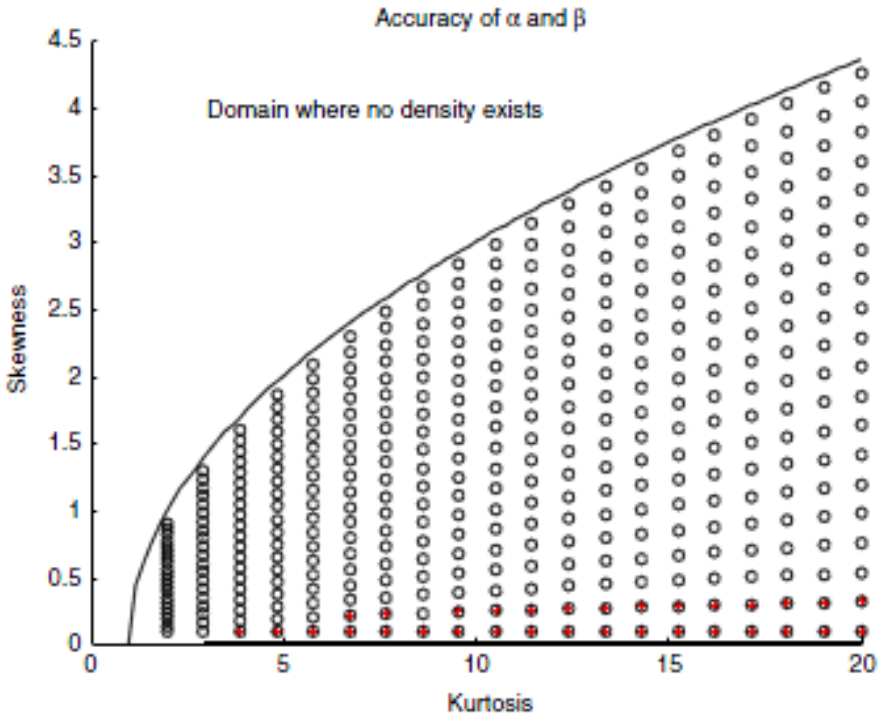


Figure 5-1. Skewness and kurtosis domain of MEM (Rockinger and Jondeau (2002).

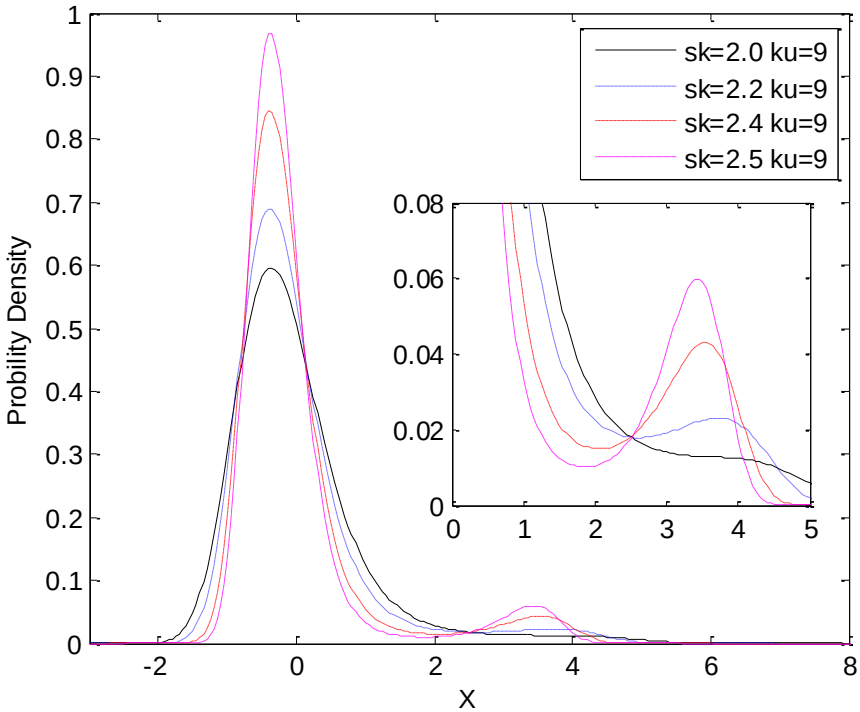


Figure 5-2. Multimodal of PDF generated by MEM in the tail region.

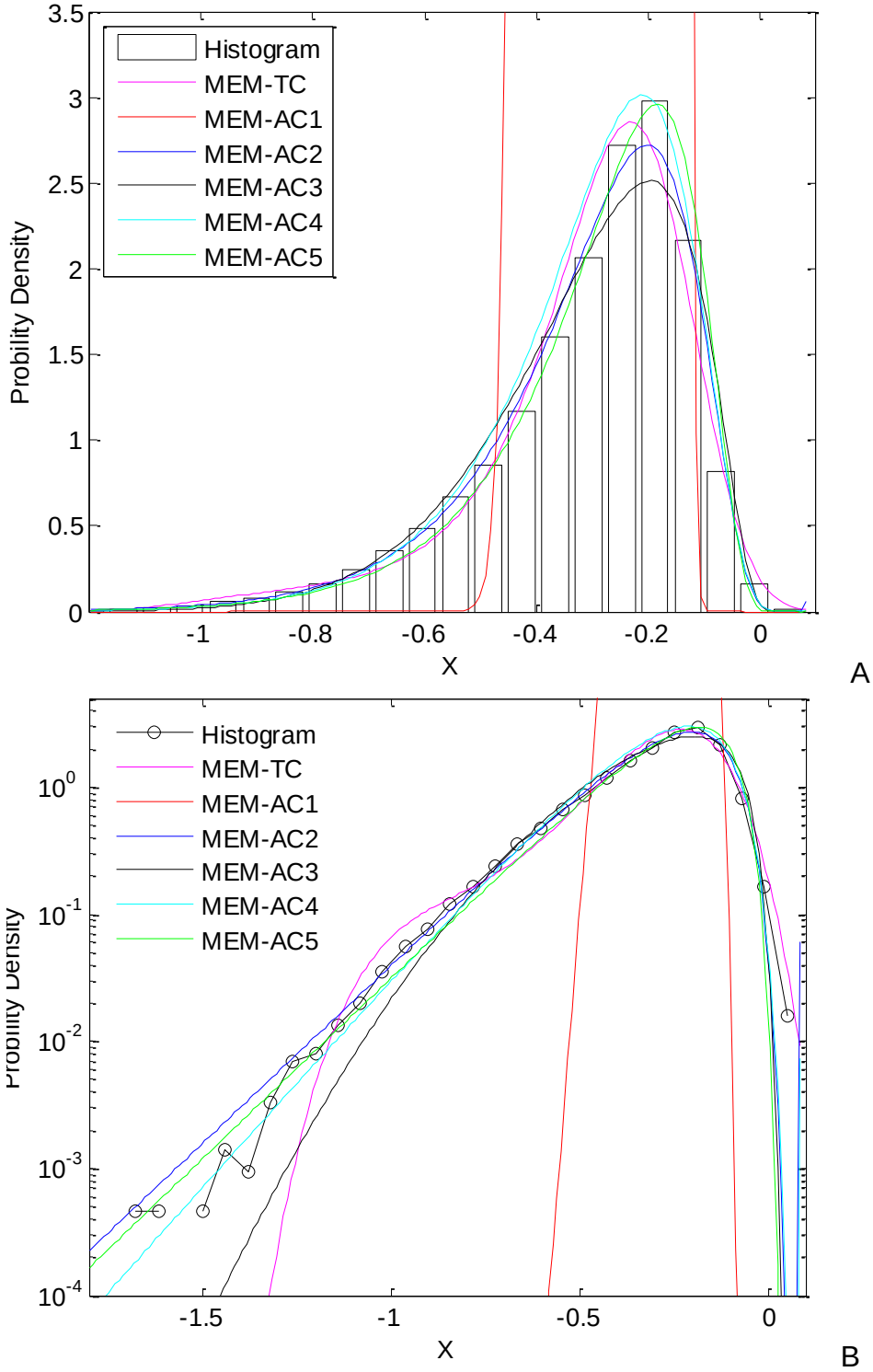


Figure 5-3. Comparison of PDF generated by MEM using traditional constrains and alternative constrains. A) linear scale, B) semi-log scale.

CHAPTER 6
HERMITE POLYNOMIAL METHOD

Traditional Formulation

An early trial of the Hermite Method is the Gram-Charlier series probability distribution (Ochi 1986). Since the distribution is given in series form, the PDF would likely have negative values at some part of its distribution, especially in the tail regions. To overcome the unwanted negative tail behavior and make the method more flexible, Winterstein used a Hermite series to functionally transform the parent Gaussian process, $U(t)$, to a Non-Gaussian process, $X(t)$ (Winterstein 1988):

$$X(t) = g[U(t)] \tag{6-1}$$

where $g(\ )$ is the functional transformation. Since the Hermite method is based on the monotonicity of CDFs of $U(t)$ and $X(t)$, function $g(\ )$ must be a monotonic function.

According to the actual characteristics of the nonlinear system or non-Gaussian excitation, it is convenient to separately consider “softening” processes (with wider and thicker tails than Gaussian distribution; e.g. $\gamma_4 > 3$) and “hardening” processes (with narrower and thinner than Gaussian tails; e.g. $\gamma_4 < 3$).

For softening process, the functional transformation, g , to a standardized process could be approximated by using the first four moments (Winterstein 1988) :

$$X = g[U] = k\left[U + c_3\left(U^2 - 1\right) + c_4\left(U^3 - 3U\right)\right] \tag{6-2}$$

$$k = \frac{1}{\sqrt{1 + 2c_3^{\ 2} + 6c_4^{\ 2}}} \tag{6-3}$$

where for convenience, X has been standardized to have zero mean value and a variance of one. k is a scaling factor that makes X have unity variance. The

coefficients c_3 and c_4 are parameters, which control the intensity of the non-Gaussian properties (e.g. skewness γ_3 and kurtosis γ_4).

The numerical relationship between γ_3, γ_4 and c_3, c_4 were approximated by Winterstein (1988) as follows:

$$c_3 = \frac{\gamma_3}{4 + 2\sqrt{1 + 1.5(\gamma_4 - 3)}} \tag{6-4a}$$

$$c_4 = \frac{\sqrt{1 + 1.5(\gamma_4 - 3)} - 1}{18} \tag{6-4b}$$

Since the accuracy of the above numerical relationship dictates the accuracy and the range of application of Hermite method, Winterstein (2000) modified his earlier formula and gave a set of equations with improved accuracy.

$$c_3 = \frac{\gamma_3}{6} \frac{1 - 0.015|\gamma_3| + 0.3\gamma_3^{\ 2}}{1 + 0.2(\gamma_4 - 3)} \tag{6-5a}$$

$$c_4 = c_{40}\left[1 - \frac{1.43\gamma_3^{\ 2}}{(\gamma_4 - 3)}\right]^{1 - 0.1\gamma_4^{\ 0.8}} \tag{6-5b}$$

$$c_{40} = \frac{[1 + 1.25(\gamma_4 - 3)]^{1/3} - 1}{10} \tag{6-5c}$$

After the relationship between c_3, c_4 and γ_3, γ_4 is established, the PDF of $X(t)$ is given by (Grigoriu 1984)

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{u^2(x)}{2}\right] \frac{du(x)}{dx} \tag{6-6}$$

where x is expressed in terms of u for softening process as in Eq(6-2). The above equation requires that Eq(6-2) be inverted and its inverse is given by (Winterstein 1987)

$$u(x)=\left[\sqrt{\xi^2(x)+c}+\xi(x)\right]^{\frac{1}{3}}-\left[\sqrt{\xi^2(x)+c}-\xi(x)\right]^{\frac{1}{3}}-a \tag{6-7}$$

where

$$\begin{aligned}&\xi(x)=1.5b\left(a+\frac{x}{k}\right)-a^3\\&a=\frac{c_3}{3c_4},b=\frac{1}{3c_4},c=\left(b-1-a^2\right)^3\end{aligned} \tag{6-8}$$

For hardening process ($\gamma_4<3$), the functional transformation could be the same form as in Eq(6-2), but U is expressed in terms of X, because we want to expand X, so that the kurtosis could be 3 after functional transformation :

$$U=g^{-1}(X)=X+c_3\left(X^2-1\right)+c_4\left(X^3-3X\right) \tag{6-9}$$

Using Eq(6-6) and Eq(6-9), the PDF for hardening process can be determined.

Limitations of the Traditional Hermite Method

The relationship between c_3, c_4 and γ_3, γ_4 in Eq(6-4) and Eq(6-5) are derived from the Taylor expansion of the expectation of Hermite series, implying that the above equations are precise for weak non-Gaussian data. When the standard Hermite method is used to model strongly non-Gaussian data (data with high skewness and/or kurtosis values), this method fails to accurately capture the non-Gaussian characteristics. Also, since the numerical relationship between c_3, c_4 and γ_3, γ_4 is approximate, it is not possible to clearly define a boundary, which the Hermite method will be ineffective.

Improved Hermite Method

To improve the standard Hermite method for softening process and get more accurate relationship between c_3, c_4 and γ_3, γ_4, we can calculate the third- and fourth

moments on both sides of Eq(6-2). It will give the following set of equations (Tognarelli, Zhao and Kareem 1997)

$$\begin{aligned}\gamma_3(c_3,c_4) &= k^3\left(8c_3{}^3 + 108c_3c_4{}^2 + 36c_3c_4 + 6c_3\right)\\ \gamma_4(c_3,c_4) &= k^4(60c_3{}^4 + 3348c_4{}^4 + 2232c_3{}^2c_4{}^2 + 60c_3{}^2\\ &\quad + 252c_4{}^2 + 1296c_4{}^3 + 576c_3{}^2c_4 + 24c_4 + 3)\end{aligned} \tag{6-10}$$

A major challenge in the Hermite method is to solve Eq(6-10), which is a coupled pair of non-linear equations that can be iteratively solved. The numerically solved relationship between c_3, c_4 and γ_3, γ_4 are shown in Figure 6-1 and Figure 6-2. From the two figures, we can see that the surface of c_3 is anti-symmetric about axis $skewness = 0$ and the surface of c_4 is symmetric about surface $skewness = 0$. The region within which $\gamma_3(c_3,c_4)$ and $\gamma_4(c_3,c_4)$ have corresponding values is the practical effective region, where the HPM could generate admissible PDFs. It is also observed that the effective domain of $\gamma_3(c_3,c_4)$ and $\gamma_4(c_3,c_4)$ is symmetric.

To ensure that the function g is monotonic, the derivative of g with respect to U must be larger than zero

$$\frac{dg}{du} = k\left[1 + 2c_3U + c_4\left(3U^3 - 3\right)\right] > 0 \tag{6-11}$$

Eq(6-11) must be valid for all U. This requirement gives

$$c_3{}^2 + 3c_4(3c_4 - 1) < 0 \tag{6-12}$$

Eq(6-12) actually represents a elliptic region of c_3 and c_4

$$\left(\frac{c_3}{\frac{1}{2}}\right)^2 + \left(\frac{c_4 - \frac{1}{6}}{\frac{1}{6}}\right)^2 < 1 \tag{6-13}$$

let

$$\begin{cases} c_3 = \frac{1}{2} \cdot \sin(t) \\ c_4 = \frac{1}{6}\cos(t) + \frac{1}{6} \end{cases} \tag{6-14}$$

where t is parameter that relates c_3 and c_4.

By substituting points of c_3 and c_4 on the border of the ellipse in Eq(6-14) into Eq (6-10), the analytical effective region of skewness and kurtosis could be established by mapping the $c_3 - c_4$ points to $sk - kt$ points. Figure 6-3 illustrates the resulting effective region for c_3 and c_4 overlaid upon the analytically effective region for skewness and kurtosis. The 'practical border' in Figure 6-3B represents to the region where solutions to Eq(6-10) could result in admissible PDFs (no negative or complex values). It is observed that practical border almost exactly follow the analytically border, indicating the correctness of analytical border. The HPM may be considered as a model for any random variable whose skewness and kurtosis falls within the practical region.

To improve the accuracy for hardening process, the functional transformation could be the same form as in Eq(6-2). We do not need to invert Eq(6-11), because U has already been expressed in terms of X .But we still have find the relationship between c_3, c_4 and γ_3, γ_4, which could be found by setting

$$\begin{aligned} \gamma_3 &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{u^2(x)}{2}\right] \frac{du(x)}{x} \cdot x^3 dx \\ \gamma_4 &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{u^2(x)}{2}\right] \frac{du(x)}{x} \cdot x^4 dx \end{aligned} \tag{6-15}$$

In the above equations, c_3 and c_4 are implicitly included in the pair of coupled nonlinear equations, which can also be iteratively solved until calculated c_3 and c_4 generate exact target skewness and kurtosis as what we did for softening process. If for higher efficiency, surface fitting to numerical c_3 and c_4 could also be conducted. Since the non-Gaussian wind pressure data have kurtosis larger than 3, this thesis will focus on softening process ($\gamma_4 > 3$).

Surface-fitting Algorithm

As mentioned earlier, a major challenge in the Hermite method is obtaining a solution to the coupled non-linear equations in Eq (6-10). They can be iteratively solved to get the shape parameters in terms of skewness and kurtosis. However, when Hermite method is used for practical application, the numerical iteration is troublesome for programming and time-consuming. For simplicity and efficiency, a surface fitting approach for c_3 and c_4 can be conducted. The surface fitting was conducted in surface-fitting tool box of Matlab R2009B. The tool box can be used to interactively fit different surfaces to the data and then view the plots in the Surface Fitting Tool's graphical user interface. Figure 6-4 and Figure 6-5 show the surface-fitting process.

Thus, in this way, a polynomial function gives an appropriate fitting of the numerical relationships, $c_3 = f_1(sk, kt)$ and $c_4 = f_2(sk, kt)$. In Figure 6-4 for the surface fitting of c_3, the order of skewness is 5 and the order of kurtosis is 4. For c_4, the order of skewness is 4 and kurtosis is 5. The surface fits developed in the current study are given by Eqs (6-16&17):

$$c_3 = p1\cdot\gamma_3 + p2\cdot\gamma_3\gamma_4 + p3\cdot\gamma_3^{\ 3} + p4\cdot\gamma_3\gamma_4^{\ 2} + p5\cdot\gamma_3^{\ 3}\gamma_4 + p6\cdot\gamma_3\gamma_4^{\ 3} + p7\cdot\gamma_3^{\ 5} + p8\cdot\gamma_3^{\ 3}\gamma_4^{\ 2} + p9\cdot\gamma_3\gamma_4^{\ 4} \tag{6-16}$$

$$c_4 = p1 + p2\cdot\gamma_4 + p3\cdot\gamma_3^{\ 2} + p4\cdot\gamma_4^{\ 2} + p5\cdot\gamma_3^{\ 2}\gamma_4 + p6\cdot\gamma_4^{\ 3} + p7\cdot\gamma_3^{\ 4} + p8\cdot\gamma_3^{\ 2}\gamma_4^{\ 2} + p9\cdot\gamma_4^{\ 4} + p10\cdot\gamma_3^{\ 4}\gamma_4 + p11\cdot\gamma_3^{\ 2}\gamma_4^{\ 3} + p12\cdot\gamma_4^{\ 5} \tag{6-17}$$

The coefficients are given in Table 6-1.

By looking at Figure 6-4&5, the surfaces generated for c_3 and c_4 and closely track the target points.

Test statistics are determined to evaluate the goodness of fit of the fitted surface. The accuracy of surface fits for c_3 and c_4 is demonstrated in Table 6-2.

To show the improved accuracy of proposed surface fitting formula, let $\gamma_3^{\ *}$ and $\gamma_4^{\ *}$ correspond to the theoretical target values and let γ_3 and γ_4 correspond to the calculated values using the above three set of fitting formulas (Winterstein's fitting formula Version I, Version II and proposed surface fitting) . Percentage error can be defined as

$$\Delta\alpha = \frac{\sqrt{\left(\gamma_3 - \gamma_3^{\ *}\right)^2 + \left(\gamma_4 - \gamma_4^{\ *}\right)^2}}{\sqrt{\left(\gamma_3^{\ *}\right)^2 + \left(\gamma_4^{\ *}\right)^2}}\cdot 100\% \tag{6-18}$$

The percentage errors of the fitting formulas are presented in Figure 6-6, 6-7 and 6-8. In these plots, it is observed that the error caused by proposed surface fitting is within 2% for most part of the effective region, and only some larger errors (near 6%) lie at part of the edge with kurtosis around 3 (very week non-Gaussian effect). Comparing the errors caused by Winterstein's fitting Version I and Version II, the accuracy of Version II has improved significantly. However, the fitting formula Version I and Version

II are a based on data with smaller deviations from Gaussian than the current study (i.e. only a small subset of the region in Figure 6-3 was under consideration). As will be demonstrated later, the dataset of wind pressure on a bluff body produces a wide range of skewness and kurtosis pairs. The existing closed form approximate solution for shape parameters was found to be inappropriate for much of the pressure data under consideration. By the above comparison, the surface fitting solution very accurately replicates the target third and fourth moment statistics.

Estimating Skewness and Kurtosis by Maximum Likelihood Method

The HPM PDF is determined as a function of the first four moments, i.e. mean, standard deviation, skewness and kurtosis. To get the four parameters, the moment method (MM) and the maximum likelihood method (MLM) could be used. As demonstrated in section 3.4, a substantial amount of statistical uncertainty is embedded in the parameter estimations when the method of moments is used. However, since the method of maximum likelihood is less sensitive to rare extreme outcomes than moment method, maximum likelihood method could serve as an alternative to the method of moments to estimate the parameters of HPM.

The process of using MLM to estimate parameters is:

1). calculate the mean, standard deviation, skewness and kurtosis as preliminary statistical estimates;

2). normalize the data using $\overline{X} = \frac{X - \mu}{\sigma}$. Since the uncertainty of mean and standard deviation is much smaller than that of skewness and kurtosis, only skewness and kurtosis will be optimize;

3). use optimization algorithm to find the optimum value of skewness and kurtosis, which would maximize the likelihood estimate, $lik(sk,kt)=\prod_{i=1}^{N} f\left(X_i \middle| (sk,kt)\right)$;

4). substitute the estimated statistics into HPM model to generate the desired PDF.

In chapter 7 and Appendix B, it will be shown that MLM is rather effective in improving the accuracy and flexibility of HPM.

Examples of Hermite Method

In this section, two examples are presented to show the performance of the Hermite probability model. Example one uses the wind tunnel pressure time history of Tap 133 with azimuth=45^0. The PDF generated by numerically solving Eq(6-10) is denoted by 'Herm Solve'. PDF generated by the proposed surface fitting is denoted by 'Herm SurFit'. 'Herm Win1' and 'Herm Win2' represent the PDFs generated by Winterstein's fitting formula version 1 (Winterstein 1988) and version 2 (Winterstein and Kashef 2000), respectively. The calculated skewness and kurtosis from the data are -1.34 and 7.19, respectively.

Figure 6-9 presents PDFs generated by Hermite models for Tap 133. From Figure 6-9, it is observed Herm Solve, Herm SurFit and Herm Win2 can closely capture the shape of data histogram and the three PDFs are close to each other, indicating the approximated relationship between c_3, c_4 and sk, kt in Herm SurFit and Herm Win2 is accurate.

The accuracy of Herm SurFit and Herm Win2 is also confirmed in Table 6-3. From Table 6-3, we can see that Herm Solve consumes the longest time among the four candidates, but, as a result of the simplicity of HPM, Herm Solve is still efficient, capable

of generating 62 PDFs per second. The computation time for the other three models is almost the same. Only ‘Herm Win1’ generates considerable error for kurtosis. Winterstein’s fitting formula version 2 is precise for this case.

The next example uses the pressure time history of Tap 236 with azimuth=45^0. This time history is chosen to show performances of the four models in cases of strong Non-Gaussian properties. The results are presented in Figure 6-10. In this figure, it is noted that Herm Win1 and Herm Win2 are more deviated than Herm Solve and Herm SurFit.

When the pressure Tap 236 experiences stronger Non-Gaussian effects, the performances of the four Hermite algorithms are studied. The measured skewness and kurtosis from the data are -1.97 and 11.84, respectively. In this case, Table 6-4 shows that ‘Herm SurFit’ uses the leas time, while ‘Herm Solve’ uses the longest time. By comparing the error of the four methods, we can see that ‘Herm SurFit’ is not only efficient but also precise. It is worth to note that the error of kurtosis of ‘Herm Win1’ is 48%, making the generated PDF deviate from the histogram in the mean region. Also the error of skewness and kurtosis of ‘Herm Win2’ is larger than 10%, indicating that Winterstein’s fit is not appropriate for strong Non-Gaussian data.

Based on the discussion of the accuracy and efficiency of HPM, the HPM using numerical solving algorithm or proposed surface fitting algorithm is robust and flexible for the practical application to modeling the PDFs of non-Gaussian data.

Table 6-1. Coefficients in Eqs(6-16&17)

Coefficients	For c_3	For c_4
P1	0.1967	-0.0721
P2	-1.646e-2	0.03176
P3	1.809e-2	-0.02942
P4	7.438e-4	-0.00179
P5	-9.209e-4	0.002348
P6	-1.366e-5	5.965e-5
P7	1.527e-4	-6.282e-4
P8	1.07e-5	-6.355e-5
P9	8.823e-8	-9.692e-7
P10	--	1.497e-5
P11	--	5.457e-7
P12	--	6.049e-9

Table 6-2. Accuracy of surface fitting

Fit name	Fit type	SSE	R-squared	Adj-Rsq	RMSE	# Coeff
Surface fitting for c_3	Ploy54	0.0277	0.9988	0.9987	0.0071	20
Surface fitting for c_4	Ploy45	0.0012	0.9994	0.9994	0.0017	20

Footnote: SSE is the sum of squares due to error of the fit. A value closer to zero indicates a fit that is more useful for prediction. R-square is the square of the correlation between the response values and the predicted response values. A value closer to 1 indicates that a greater proportion of variance is accounted for by the model. Adj R-sq is the degrees of freedom adjusted R-square. A value closer to 1 indicates a better fit. RMSE is the root mean squared error or standard error. A value closer to 0 indicates a fit that is more useful for prediction. # Coeff is the number of coefficients in the model.

Table 6-3. Comparison of Hermite models using tap 133

Hermite models	Computation time	Generated skewness	Generated kurtosis	Absolute error of skewness (%)	Absolute error of kurtosis (%)
Herm solve	0.0160	-1.34	7.19	0	0
Herm surfit	0.0060	-1.35	7.22	1.5	0.4
Herm win1	0.0050	-1.31	9.69	3.0	35
Herm win2	0.0050	-1.37	7.17	3.0	2.8

Table 6-4. Comparison of Hermite models using tap 236

Hermite models	Computation time	Generated skewness	Generated kurtosis	Absolute error of skewness (%)	Absolute error of kurtosis (%)
Herm solve	0.0150	-1.97	11.84	0.0	0.0
Herm surfit	0.0040	-2.00	11.81	1.5	0.0
Herm win1	0.0050	-1.86	17.52	5.6	48
Herm win2	0.0050	-2.17	13.10	10.1	10.6

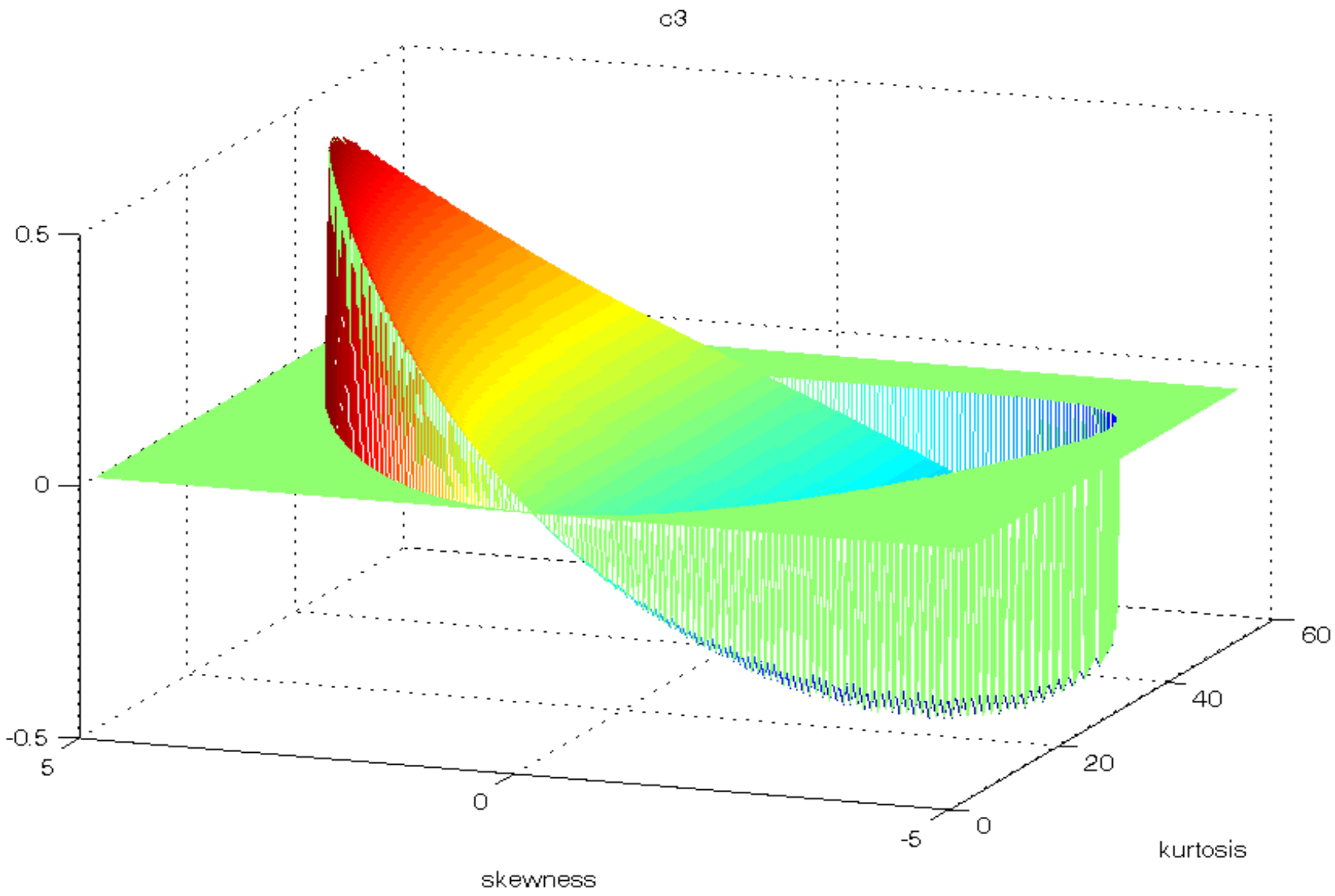


Figure 6-1. Numerically solved relationship between c_3 and γ_3, γ_4.

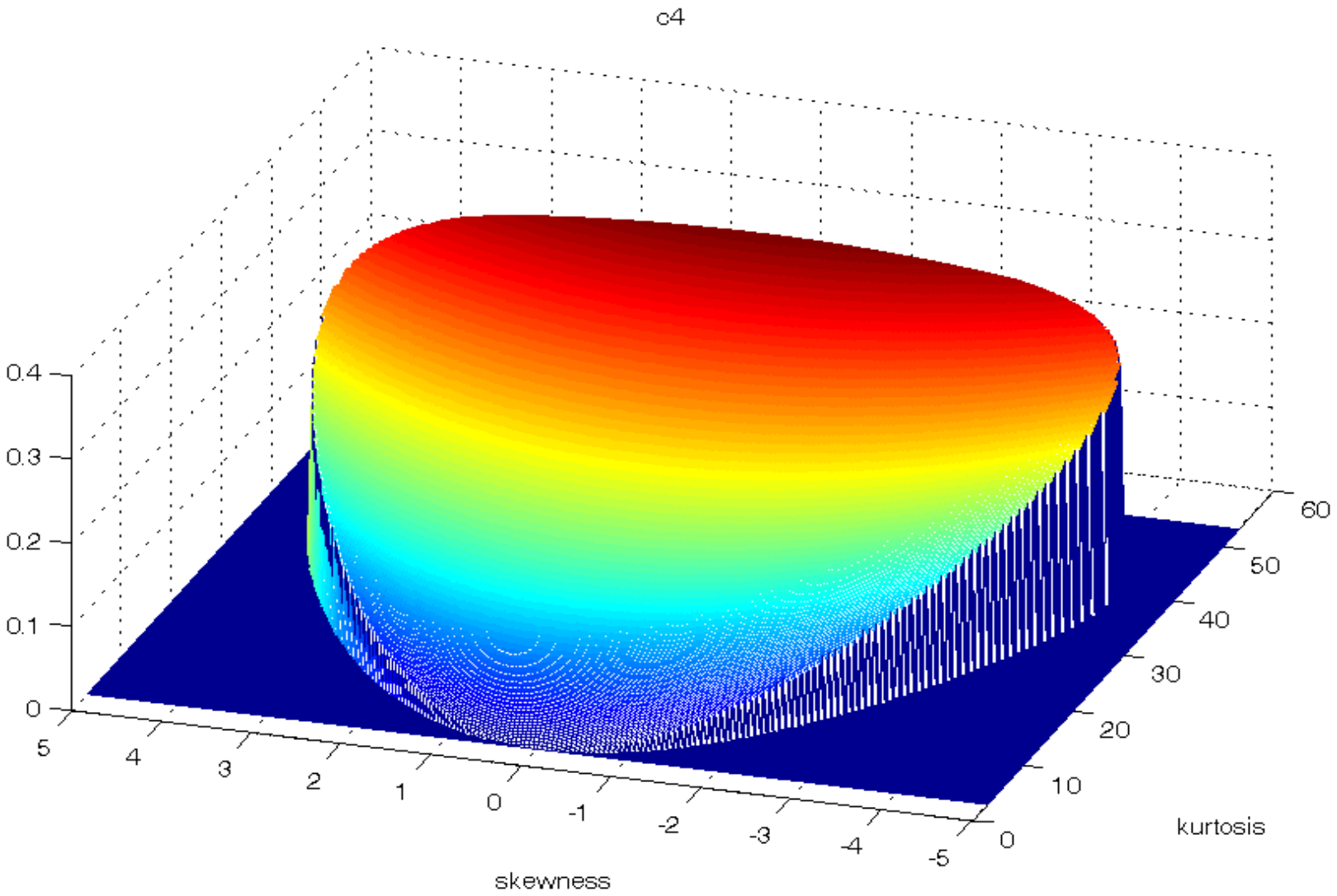


Figure 6-2. Numerically solved relationship between c_4 and γ_3, γ_4.

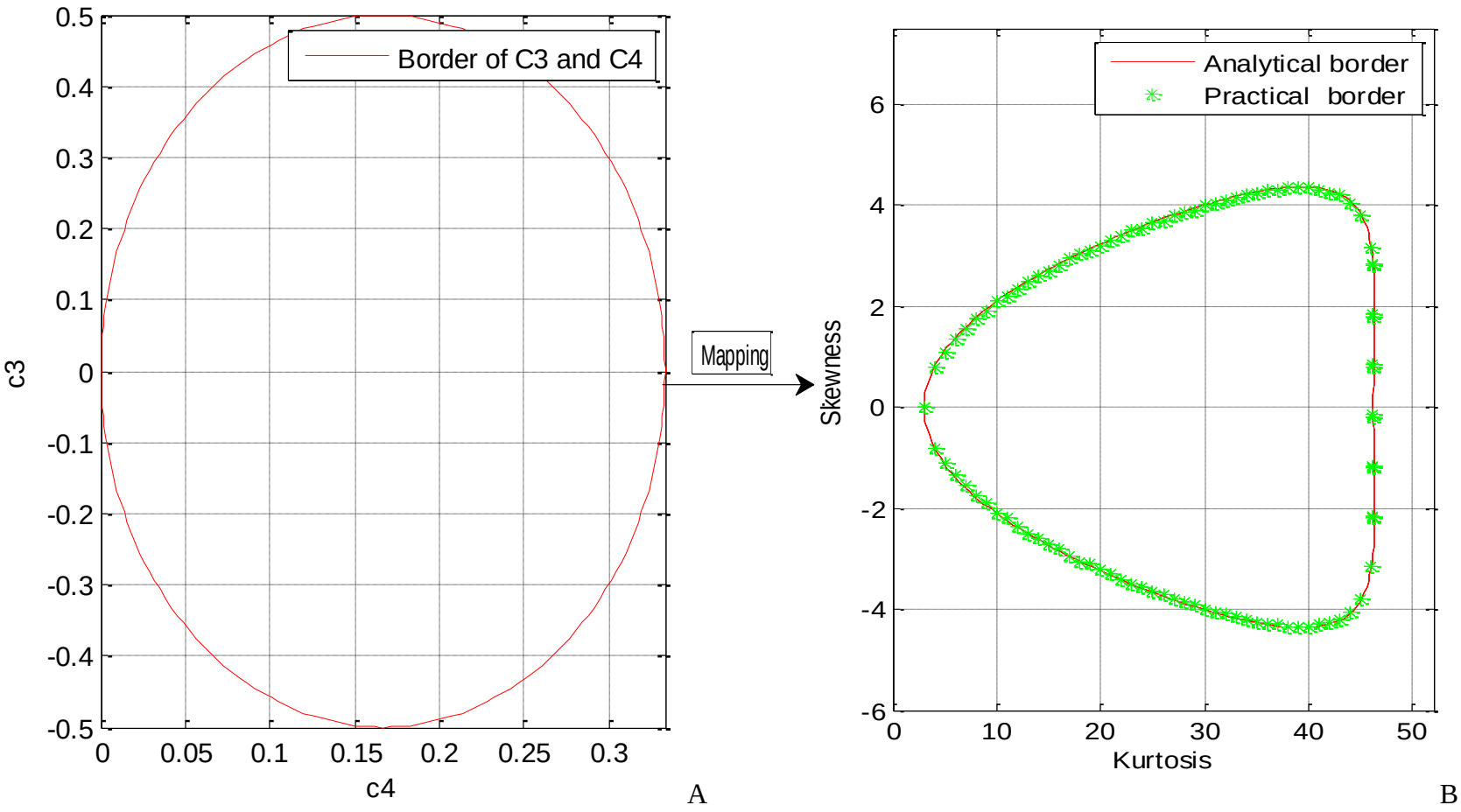


Figure 6-3. Mappring effective region of c3 and c4 to effective of sk and kt. A) effective region of c3 and c4, B) effective region of sk and kt.

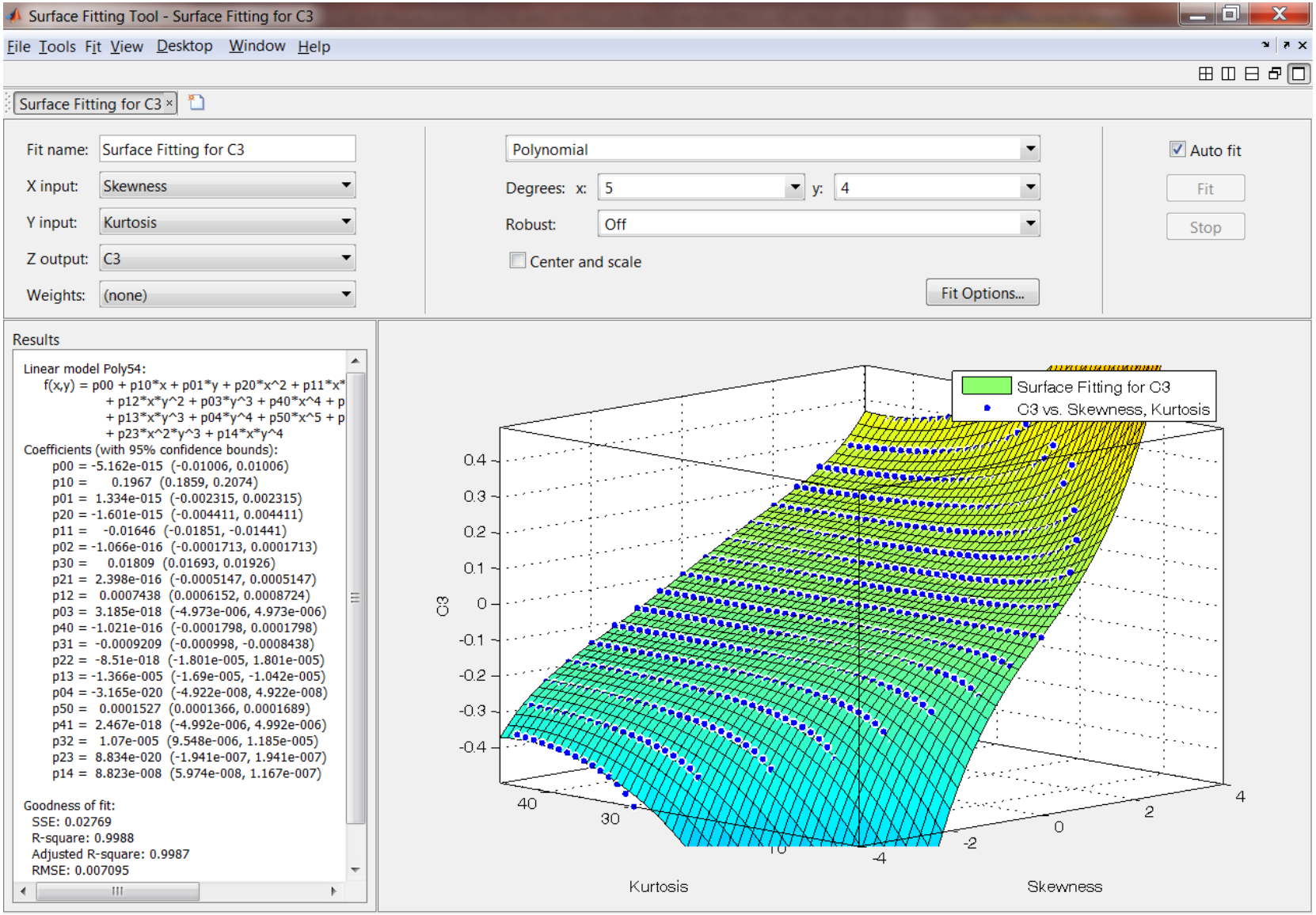


Figure 6-4. Surface fitting for c_3

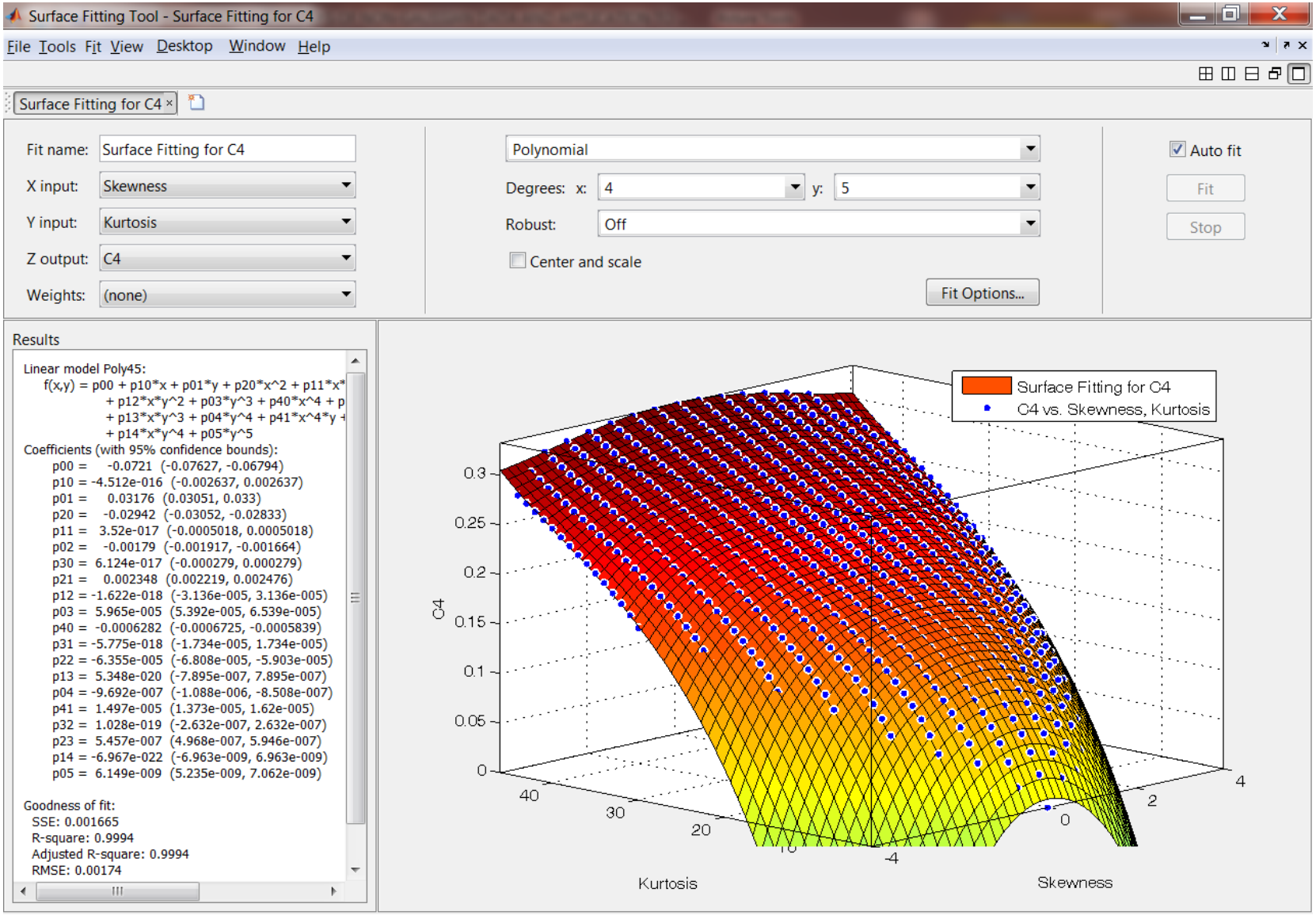
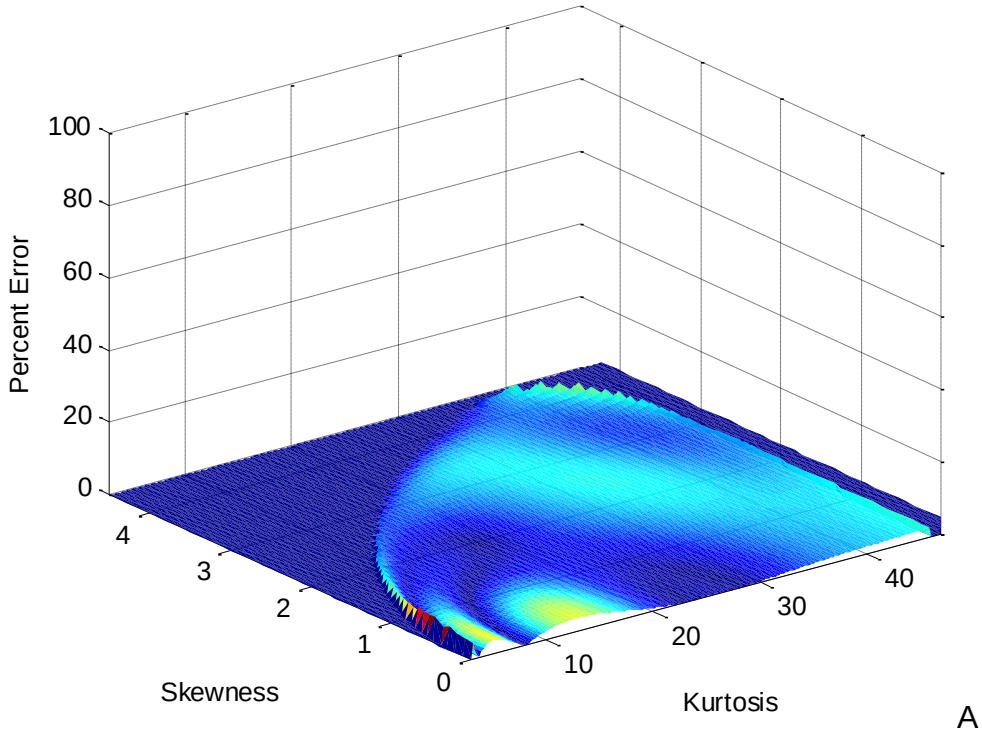


Figure 6-5. Surface fitting for c_4



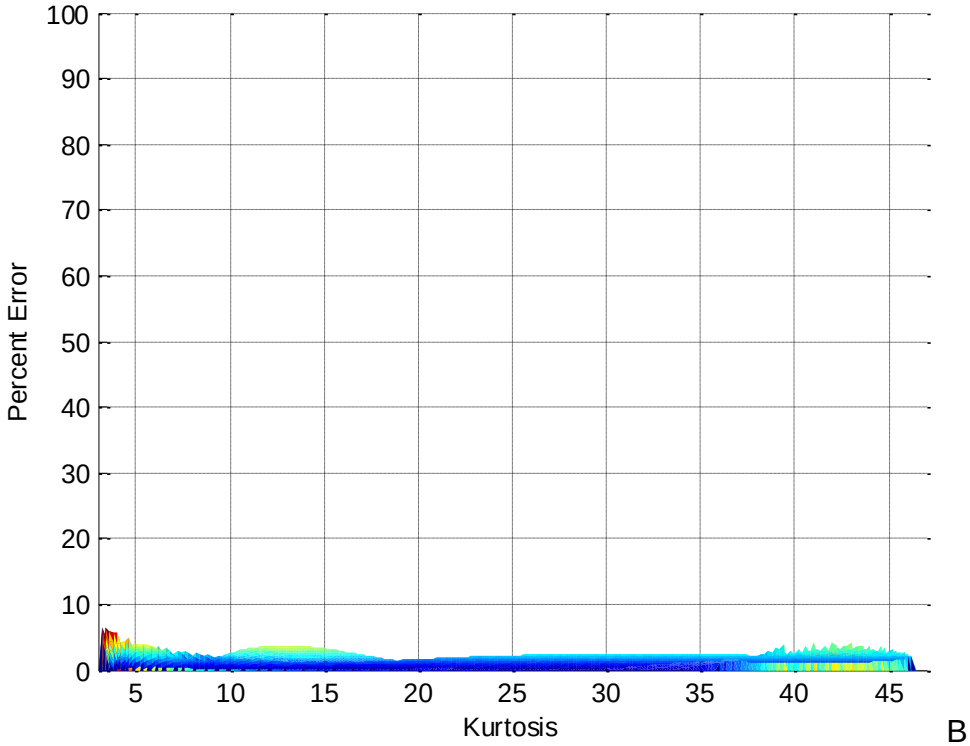


Figure 6-6. Percent error of proposed surface fitting: A) Orthogonal view, B) Side view.

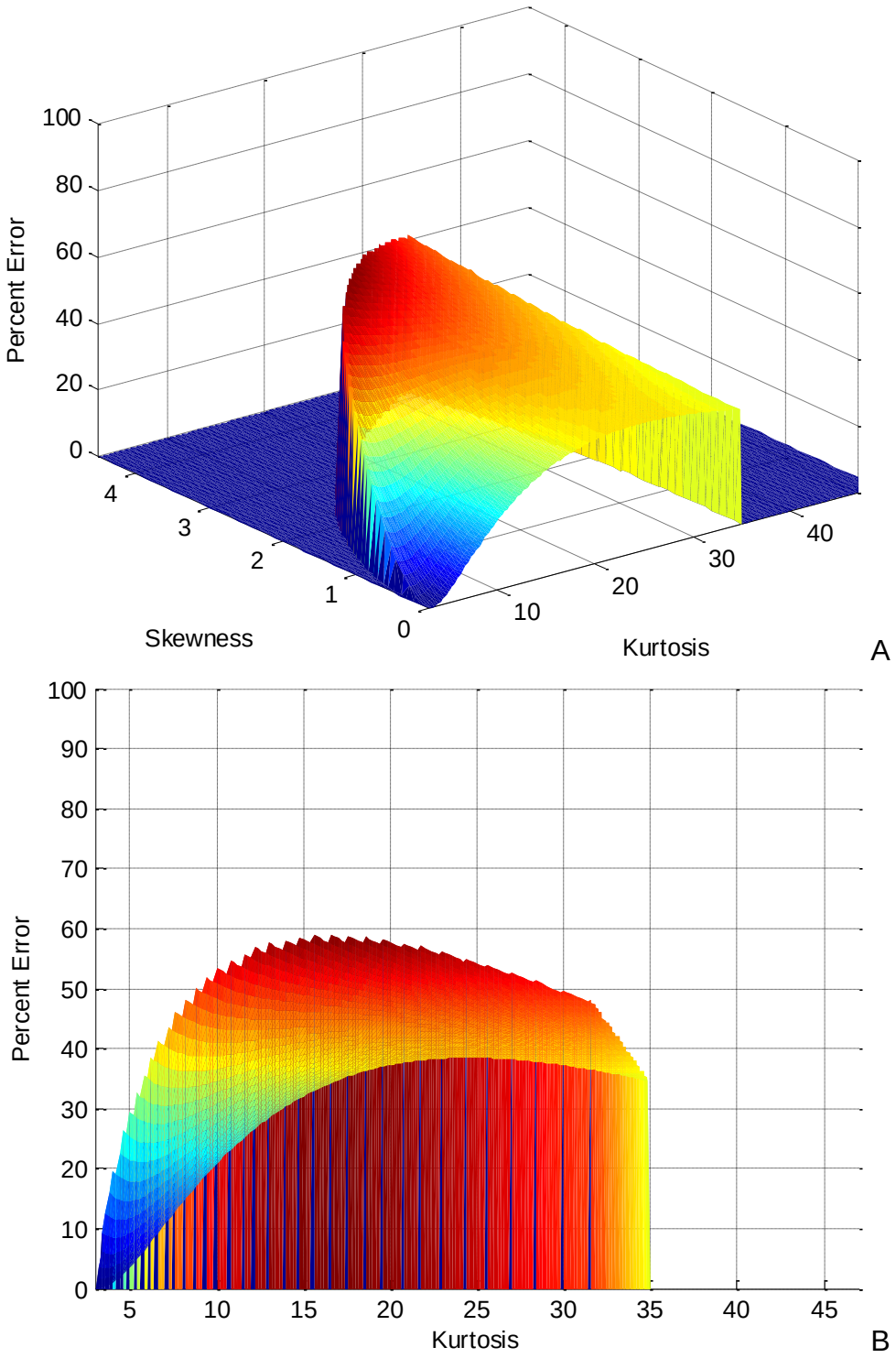


Figure 6-7. Percent error of Wintestein's fitting Version I: A) Orthogonal view. B) Side view.

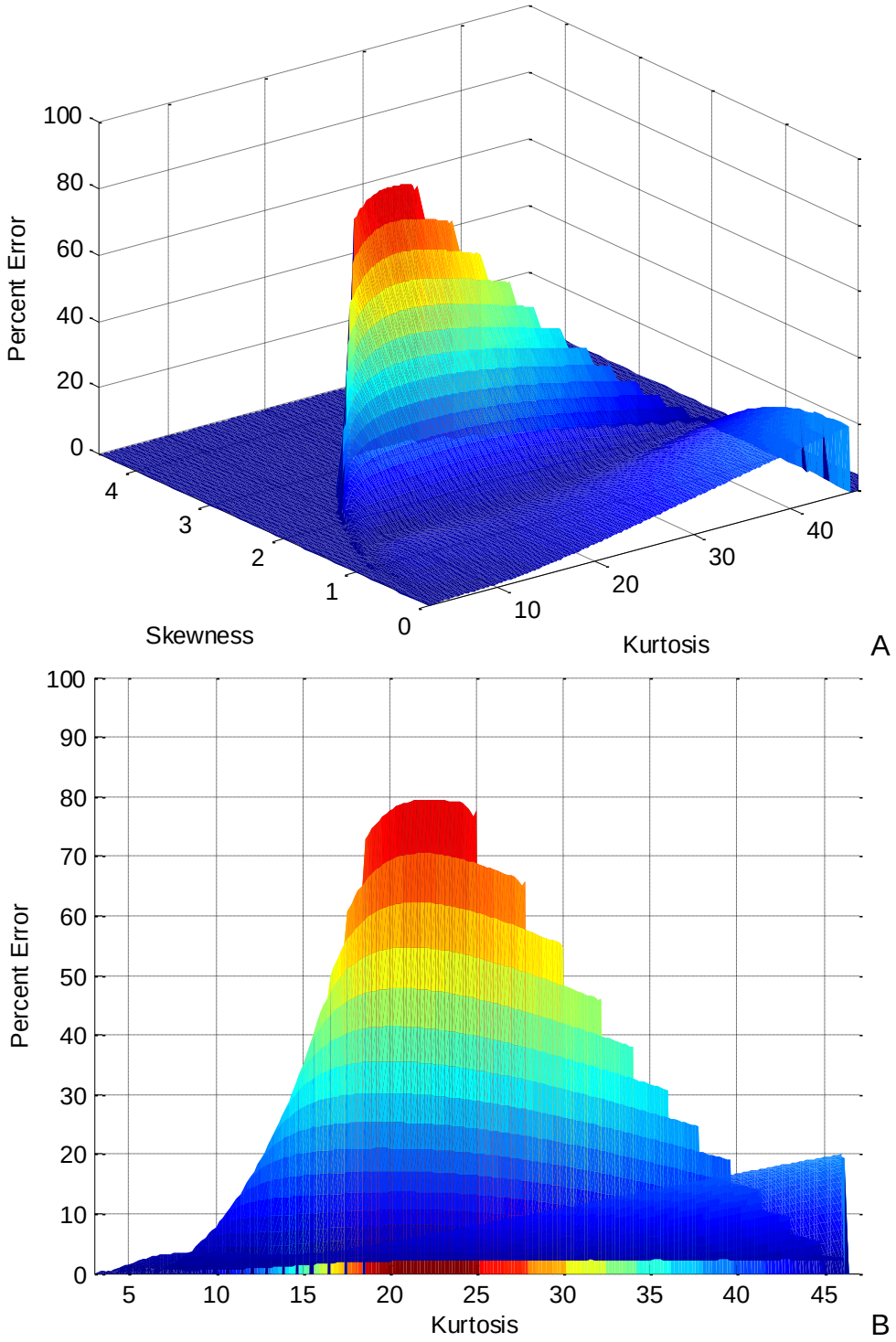


Figure 6-8. Percent error of Winterstein's fitting Version II: A) Orthogonal view. B) Side view.

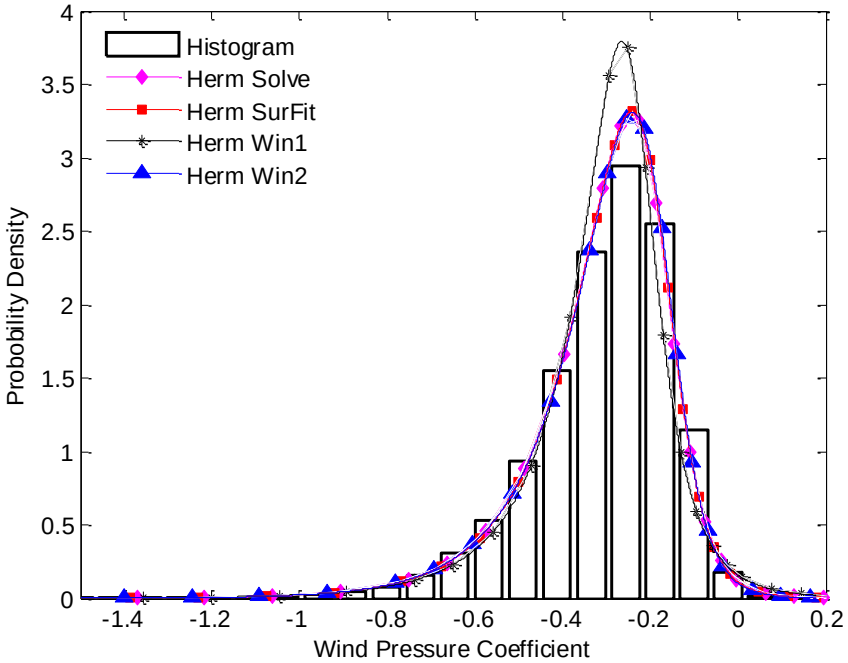


Figure 6-9. Comparison of Hermite PDFs generated by different solving algorithm to Eq(6-10).

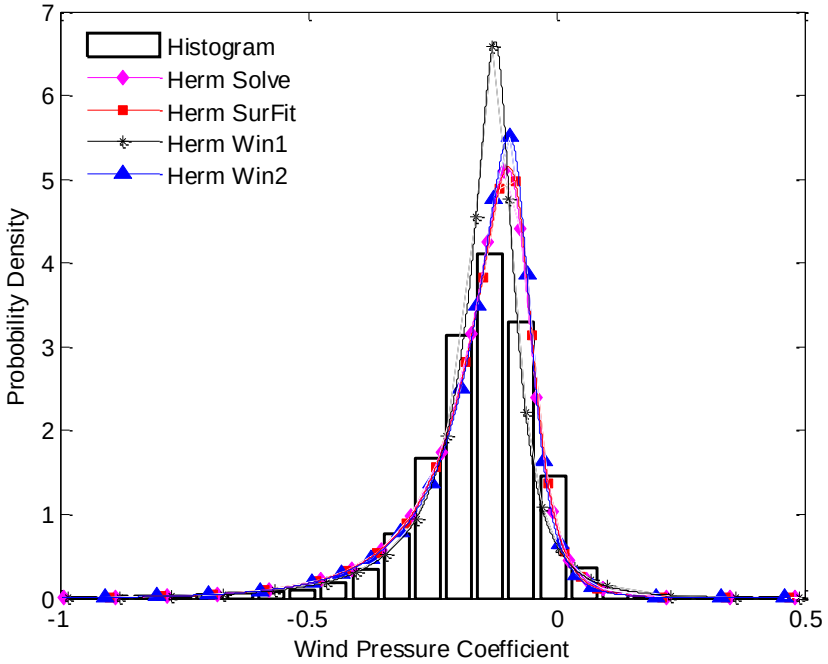


Figure 6-10. Comparison of Hermite PDFs generated by different solving algorithm to Eq(6-10).

CHAPTER 7
APPLICATION AND COMPARISON OF PDF MODELS

Test of Goodness of Fit (GOF)

Pressure time series are fitted with a set of candidate families of distributions, such as Gaussian, Shifted Lognormal, Shifted Gamma, GEV, Weibull, Rayleigh and Exponential distribution. In this fitting process, maximum likelihood method is applied to estimate the parameters for each candidate family of distribution. Of the common candidate families of distributions, Gaussian, Shifted Lognormal, Shifted Gamma and GEV have been studied and demonstrated their applicability for certain circumstances. Thus, the four common PDF models will be chosen and compared to HPM. For HPM, PDFs using parameters estimated by the method of moments (HPM-MM) and the maximum likelihood method (HPM-ML) are compared with PDFs generated from the four common PDF models.

Based on visual inspection and previous researches, the six PDF candidates are chosen for the test of GOF. These candidate families are Gaussian, Shifted Lognormal, Shifted Gamma, GEV, Hermite-MM and Hermite-ML. Three standard tests of GOF were used: mean square error (MSE), mean percentage error (MPE) and the Kolmogorov-Smirnov (KS) test. The KS test essentially compares the maximum error between the model and exact cumulative distribution functions. The test statistic is defined as (Massey 1951)

$$D = \max_{1 \le i \le N} \left(F(Y_i) - \frac{i-1}{N}, \frac{i}{N} - F(Y_i) \right) \tag{7-1}$$

where F is the cumulative distribution to be tested which must be a continuous distribution, N is number of data points.

Since the MSE, MPE and KS tests heavily favor the mean region of the distributions where the probability weight is highest, these tests are less sensitive to the tail region. Given the dependence of damage initiation and propagation on the extreme uplift peaks, it is important to accurately model this tail region of the distribution where the probability weight is low but critical (Cope et al (2005). So it is necessary to use a second kind GOF test that focuses on the tail region. This can be achieved by Anderson-Darling (AD) test of GOF. In contrast to KS test, AD test is a “quadratic” test as shown in Eq 7-2, because it is based on a weighted square of the vertical distance between the empirical and fitted CDFs (Anderson and Darling 1954)

$$A^2 = -N - \sum_{i=1}^{N} \frac{(2i-1)}{N} \{\ln[F(Y_i)] + \ln[1 - F(Y_{N+1-i})]\} \tag{7-2}$$

where F is the theoretical cumulative distribution of the distribution being tested which must be a continuous distribution, N is number of data points. For all the above four tests of GOF, lower values indicate a better fit.

Distribution of Skewness and Kurtosis of Wind Pressure

Figure7-1 shows the effective region for the Hermite Polynomial Method (HPM) versus the distributions of skewness and kurtosis for wind flows, parallel, cornering and perpendicular to the ridge of the building model. The distributions of skewness and kurtosis lie in a region close to the border of effective region of HPM, indicating that when skewness is large, kurtosis tends to be large. From figure 7A, it is observed that the skewness for parallel wind ranges from -1.77 to -0.06, and kurtosis from 3.35 to 8.33. And all the sk-kt points lie in the effective region for parallel winds. For cornering winds (Figure 7B), some points with very high skewness and kurtosis, i.e. $skewness = -3.14$ and $kurtosis = 20.00$, are observed. And a small portion of sk-kt

points lie out of the effective region. Since these points lying out of effective region are very close to the border, HPM may still be applicable, using MLM with initial values in the effective region. For perpendicular winds, certain amount of points with positive skewness is observed and all the points lie in the effective region. By comparing Figure 7 A, B and C, it is observed that the non-Gaussian effects of cornering winds are strongest among the three wind directions. So when comparing the goodness of fit in Chapter 7, the cornering wind will be studied in detail.

For full-scale data, the distribution of skewness and kurtosis is plotted in Figure 7-1D. It is also observed that the sk-kt points lie in the effective region and the distribution has similar pattern as that of cornering winds. Generally, the four plots show high potential that HPM could serve as a flexible and robust PDF model for the pressure date in the available datasets.

Data from the taps identified in these plots are used in the examples that follow.

Comparison of PDF Models

Based on discussion in section 7.2, the majority of points in skewness and kurtosis plane lie in the region of $skewness \in (-1.8, 0.5)$ and $kurtosis \in (3, 10)$ which is within the effective region. For parallel and perpendicular winds, all the points concentrate into a much smaller sub-region. However, for cornering winds, a small proportion of the sk-kt lie out of the effective region and the sk-kt points are more scattered.

To demonstrate the effectiveness of HPM, this section presents six examples, four of which are time series from the dataset of cornering winds, the other two from full scale data. The example taps (their locations on the roof are shown in Figure 3-2) are carefully chosen to represent four different situations: moderate non-Gaussian point

located in effective region; moderate non-Gaussian point located out of effective region; strong non-Gaussian point located in effective region; strongest non-Gaussian point located in effective region. The examples compare HPM-MM (whose parameters are calculated from method of moment), HPM-ML (whose parameters are estimated from maximum likelihood method) to common PDF models, including Shifted-Lognormal, Gaussian, Shifted-Gamma, and GEV PDF models.

Example One: Moderate Non-Gaussian Point Located in Effective Region

The first example uses the time series of tap 81 with wind azimuth=45°. This tap locates in interior region on the roof and experiences moderate non-Gaussian effects. The skewness and kurtosis of this sample based on MM are -1.02 and 5.26, respectively. From MLM, the skewness and kurtosis are -1.00 and 4.99, respectively. The pair of skewness and kurtosis is in the effective region of HPM. Figure 7-2A is a complete view of the results. To demonstrate the accuracy of the PDF models in the tail regions, the PDFs in Figure 7-2B are plotted in semi-log scale. By examining the two figures, one can see that HPM-MM, HPM-ML, shifted lognormal and GEV provide better fits than the other three PDF models in the mean region. More importantly, PDFs generated by HPM-MM, HPM-ML, shifted lognormal and GEV are almost indistinguishable from the histogram (the target PDF) through the whole range of data, whereas the other PDF estimates diverge rapidly in the either the left or the right tail region or both. In this case, by visual inspection, HPM-MM, HPM-ML, shifted lognormal and GEV are competitive and the difference between the four candidates are not obvious. This is because the non-Gaussian effect is not strong.

Tests of GOF were conducted to quantitatively compare the performance of the PDF models. Table 7-1 shows the results of the MSE, MPE and KS test of GOF. The

quantitative tests give consistent results with visual inspection. But HPM-MM and HPM-ML outperform the common PDF models, especially HPM-ML, showing the best performance among all the models based on the four test statistics.

Example Two: Moderate Non-Gaussian Point Located Out of Effective Region

As another example, let us consider the time series of tap 76 in the interior region on the roof. The skewness and kurtosis estimated from MM are -1.32 and 5.36, respectively. The sk-kt values of tap 76 are very close to those of tap 81. The difference between tap 76 and tap 81 is that the sk-kt point of tap 76 lies out of the effective region, while the sk-kt point of tap 81 lies in the effective region. For HPM-MM, sk-kt values from MM are not applicable to HPM, so border point of skewness and kutosis are selected as input parameters. This problem could be improved using MLM by inputting initial values of skewness and kurtosis lying within effective region and then searching for the optimum sk-kt parameters. The skewness and kurtosis estimated from MLM are -1.52 and 7.08, respectively. In Figure 7-3, the results are presented. In Figure 7-3A, HPM-ML can better capture the shape of histogram through the whole range of data. However, HPM-MM and other common PDF models deviate from the histogram in the mean region. From Figure 7-3B, HPM-MM shows the best fit in left tail region, whereas, HPM-ML and lognormal have slight deviation.

Table 7-2 shows the GOF results. The quantitative results indicate that the HPM-ML, lognormal and GEV PDFs are good fits to the wind pressure data. But HPM-ML still has smallest test statistics. This example demonstrates that maximum likelihood method greatly improves the performance of HPM and allows the application of HPM to non-Gaussian data, whose sk-kt pair lies out of effective region.

Example Three: Strong Non-Gaussian Point Located in Effective Region

The first two examples have demonstrated the accuracy of HPM for mild non-Gaussian effect (skewness is around -1 and kurtosis is around 6), where some of the common models also accurately represent the probability in the tail. In this example, the situation of strong non-Gaussian effects will be studied, using data of tap 236. For tap 236, the skewness and kurtosis from MM almost double; that is $skewness = -2.03$ and $kurtosis = 12.58$. The skewness and kurtosis values from MLM are -1.51 and 9.08, respectively.

Figure 7-4A shows that HPM-ML stays close to the histogram in the mean region. However, HPM-MM overestimates the PDF value and the other four common PDF models underestimate the distribution in the mean region. Figure 7-4B gives us a clear view of the performance of the PDF models through the whole range of data. It is observed that HPM-MM and HPM-ML are capable of tracking the target histogram through the whole range of data, while the other four common PDF models start to deviate from the target histogram at around $Cp = -0.6$. In this example, HPM-ML demonstrates significant superior performance in the mean and tail region than the other PDF models.

Tests of GOF were also conducted for tap 236 and results are shown in Table 7-3. The quantitative results show that the HPM-ML is capable of generating a good probabilistic representation of wind pressure data for strong non-Gaussian situations, while the test statistics of the other PDF models are much larger.

Example Four: Strongest Non-Gaussian Point Located in Effective Region

In this example, time series of tap 300 will be used. This tap is located in the windward corner region and experiences the strongest non-Gaussian effects among the taps on the roof, with $sk = -3.23$ and $kt = 25.29$ estimated from moments, while $sk = -1.28$ and $kt = 13.85$ estimated from MLM. In Figure 7-5A, HPM-MM has a high spike in the mean region, which is caused by the input of high kurtosis, while HPM-ML is still very close to the histogram. The high spike of HPM-MM justifies the statement that the input values of skewness and kurtosis greatly influence the shape of generated PDFs, especially when the sk-kt values are high. However, the introduction of HPM-ML is capable of reducing the discrepancy by searching the optimum sk-kt values. In the tail region, HPM-ML still can largely follow the histogram as shown in Figure 7-5B. The other four common PDF models can neither follow the histogram of the time series in the mean region nor in the tail region.

In table 7-4, the test results of MSE, MPE, KS and AD for HPM-ML are much smaller the other PDF models, further confirming the excellent performance of HPM-ML for the strongest non-Gaussian effects.

Example Five: Strong Non-Gaussian Effect for Full-scale Data

This example will demonstrate the applicability of HPM to full-scale data from the Florida Coastal Monitoring Program landfalling hurricane data collection project. The statistics of the full-scale data are $sk = -1.58$ and $kt = 9.62$ estimated from moments, while $sk = -1.23$ and $kt = 6.38$ estimated from MLM. In Figure 7-6A and B, HPM-ML demonstrates the best performance among the PDF models studied in this paper.

In table 7-5, the test results are presented. These metrics demonstrate the superior performance of the HPM-ML throughout the distribution, particularly in the tail region.

Example Six: Most Strong Non-Gaussian Effect for Full-scale Data

The last example will demonstrate the applicability of HPM to full-scale data for strongest non-Gaussian effects. The statistics of the full-scale data are $sk = -2.59$ and $kt = 21.61$ estimated from moments, while $sk = -1.78$ and $kt = 9.30$ estimated from MLM. From the skewness and kurtosis estimated from moment, it is shown that very strong non-Gaussian still exists for full-scale measurement. In examples 3, 4, 5 and 6 by comparing the sk-kt values of strong non-Gaussian effects ($kt > 10$) and the corresponding PDFs from HPM-MM and HPM-ML, it seem that HPM-ML would search for smaller sk-kt values to improve the performance of HPM with an emphasis in the mean region, which means the match in the mean region is first satisfied and then the tail region will be considered. In Figure 7-7A and B, HPM-ML demonstrates the best performance among the PDF models studied herein.

In table 7-6, the test results are presented. These metrics demonstrate the superior performance of the HPM-ML throughout the distribution, particularly in the tail region.

Application of PDF Models

Improved PDF models could be used for fatigue and reliability analysis, extreme value analysis, etc, which would involve the cumulative frequency analysis (CFA). CFA is the analysis of the frequency of occurrence of values of a phenomenon less than a reference value. Corresponding to CFA is cumulative distribution function (CDF), which

is defined as $F(x) = P(X < x)$ with which the value of a random variable X is less than or equal to a reference value x. $F(x)$ is also called the probability of non-exceedance. In its application to extreme value analysis, the CDF value $F(x)$ is usually given and then the corresponding x is derived. The accuracy of PDF model would determine the derived x value, especially in the left tail region in the context of wind engineering.

To show the future application of HPM, the studied PDF models are used to predict the reference value, x. The probability of non-exceedance is prescribed from 0.1% to 2% with an emphasis on left tail region. Then the corresponding Cps are derived from data and PDF models. The Cps derived from data are regarded as targets to be predicted by PDF models. In Figure 18, the percent error of the predicted reference values are calculated and plotted, which is defined as,

$$Err = \left| \frac{Cp_{PDF} - Cp_{data}}{Cp_{data}} \right| \cdot 100\% \tag{7-3}$$

where Cp_{data} is calculated from data and Cp_{PDF} is derive based on corresponding PDF models.

From Figure 7-8 A&B, it is observed, for moderate non-Gaussian effects (Tap 81 and Tap 76), HPM-MM, HPM-ML and Lognormal serve as a good model to predict the reference value, largely within 10%. For strong non-Gaussian effects (e.g. Tap 236 and Tap 300), the errors that are caused by common PDF models increase significantly. However, for Tap 236, HPM-MM and HPM-ML control the errors within 10%, whereas the errors of common PDF models reach 30% to 45%. Also for Tap 300, HPM-MM and HPM-ML demonstrate much better performance than common PDF models, with errors only half of those of common PDF models. Among all the PDF models, HPM-MM

demonstrates the best performance in tail region for the entire four tap in predicting the non-exceedance values. This is because HPM-MM uses the exact third and forth moments, which are very sensitive to extreme values.

Table 7-1. Test of GOF for generated PDFs at tap 81

	HPM-MM	HPM-ML	Lognormal	Gaussian	Gamma	GEV
MSE	0.0046	0.0045	0.0077	0.0844	0.0155	0.0061
MPE	0.0317	0.0317	0.0403	0.1514	0.0582	0.0361
KS	0.006	0.0043	0.0147	0.0670	0.0251	0.0088
AD	1.2111	0.4589	11.789	348.6178	39.1104	5.0418

Table 7-2. Test of GOF for generated PDFs at tap 76

	HPM-MM	HPM-ML	Lognormal	Gaussian	Gamma	GEV
MSE	0.0145	0.0041	0.0088	0.1427	0.0329	0.0077
MPE	0.0585	0.0325	0.0474	0.2119	0.096	0.0468
KS	0.0226	0.0132	0.0163	0.1021	0.047	0.0169
AD	42.2266	9.8403	18.6367	830.9461	142.0178	19.3708

Table 7-3. Test of GOF for generated PDFs at tap 236

	HPM-MM	HPM-ML	Lognormal	Gaussian	Gamma	GEV
MSE	0.0381	0.0056	0.0386	0.1185	0.0549	0.0352
MPE	0.0759	0.034	0.0897	0.1651	0.109	0.0858
KS	0.0392	0.0176	0.0494	0.099	0.0639	0.0463
AD	105.7596	19.1841	216.0767	802.3923	329.8255	191.1084

Table 7-4. Test of GOF for generated PDFs at tap 300

	HPM-MM	HPM-ML	Lognormal	Gaussian	Gamma	GEV
MSE	0.4138	0.0190	0.1061	0.1929	0.1195	0.1175
MPE	0.1714	0.0563	0.1311	0.1838	0.1395	0.1364
KS	0.1388	0.0425	0.0934	0.1478	0.1068	0.1013
AD	1106.093	87.588	649.935	1526.636	785.422	711.724

Table 7-5. Test of GOF for generated PDFs of full-scale data

	HPM-MM	HPM-ML	Lognormal	Gaussian	Gamma	GEV
MSE	0.0205	0.0067	0.0193	0.0829	0.0209	0.0133
MPE	0.0635	0.0379	0.0622	0.1393	0.0647	0.0519
KS	0.0314	0.0139	0.0360	0.0802	0.0319	0.0229
AD	17.2409	2.3651	19.4497	126.339	18.5862	8.5260

Table 7-6. Test of GOF for generated PDFs of full-scale data

	HPM-MM	HPM-ML	Lognormal	Gaussian	Gamma	GEV
MSE	0.5483	0.0119	0.0448	0.3316	0.0974	0.0261
MPE	0.2246	0.0478	0.0856	0.248	0.1291	0.0649
KS	0.1172	0.0153	0.0339	0.1125	0.0595	0.0273
AD	216.8783	3.4057	21.3999	242.6718	55.6401	10.2511

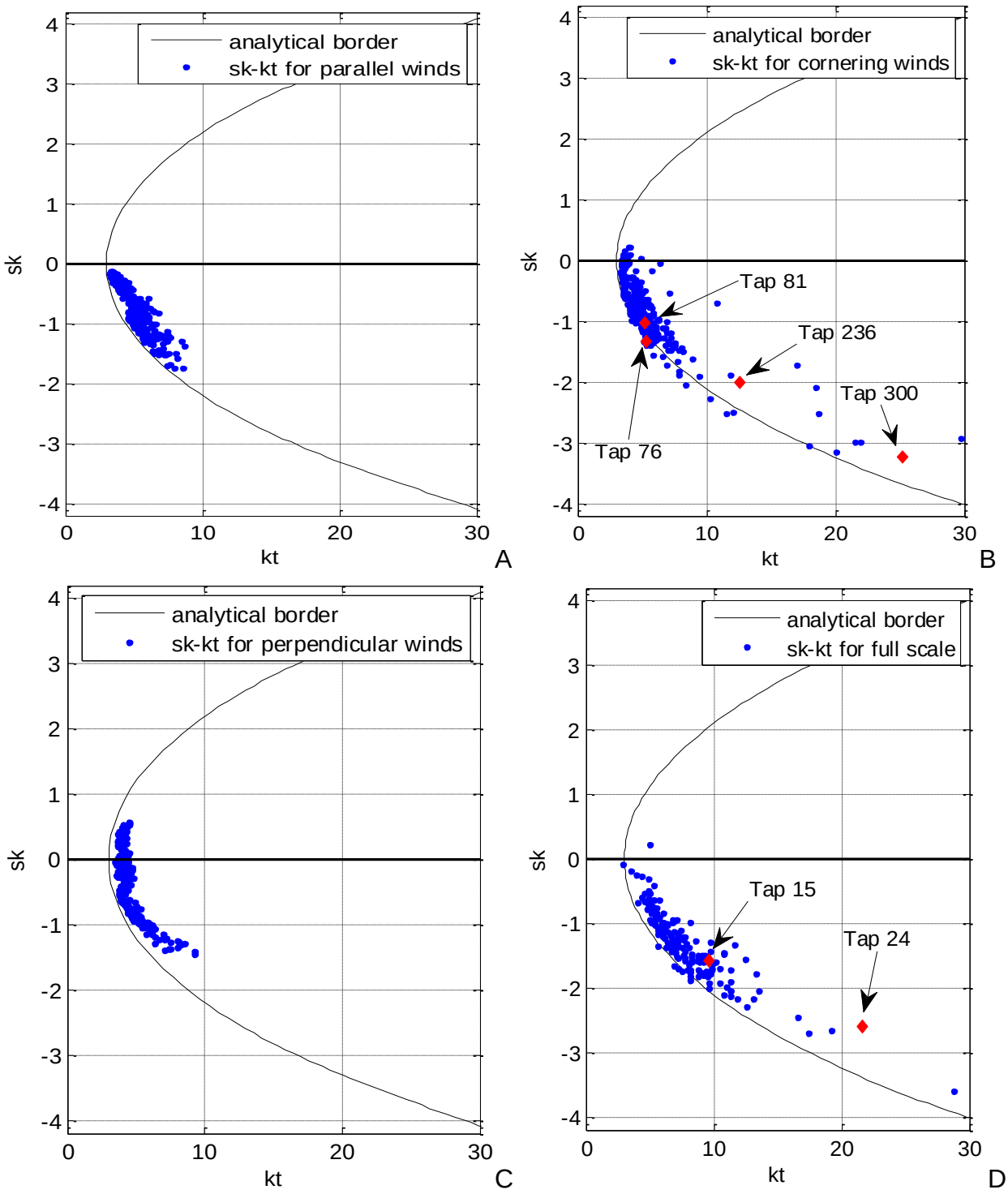


Figure 7-1. Distribution of points of skewness and kurtosis. A) parallel winds, B) cornering winds, C) perpendicular winds, D) full-scale dat.

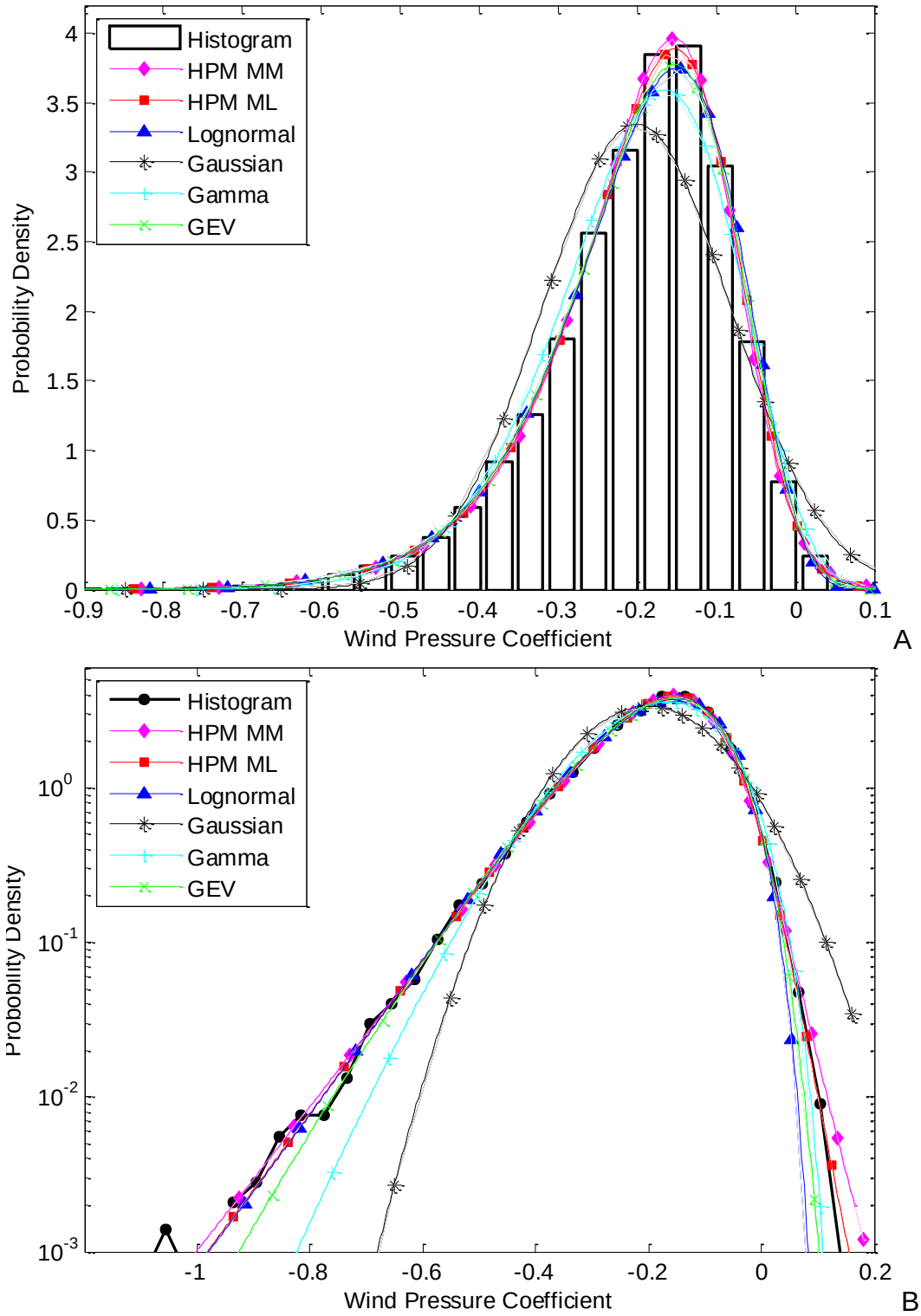


Figure 7-2.Comparison of PDFs of tap 81 (sk=-1.02, kt=5.26) generated by candidate PDF models. A) Linear scale, B) semi-log scale.

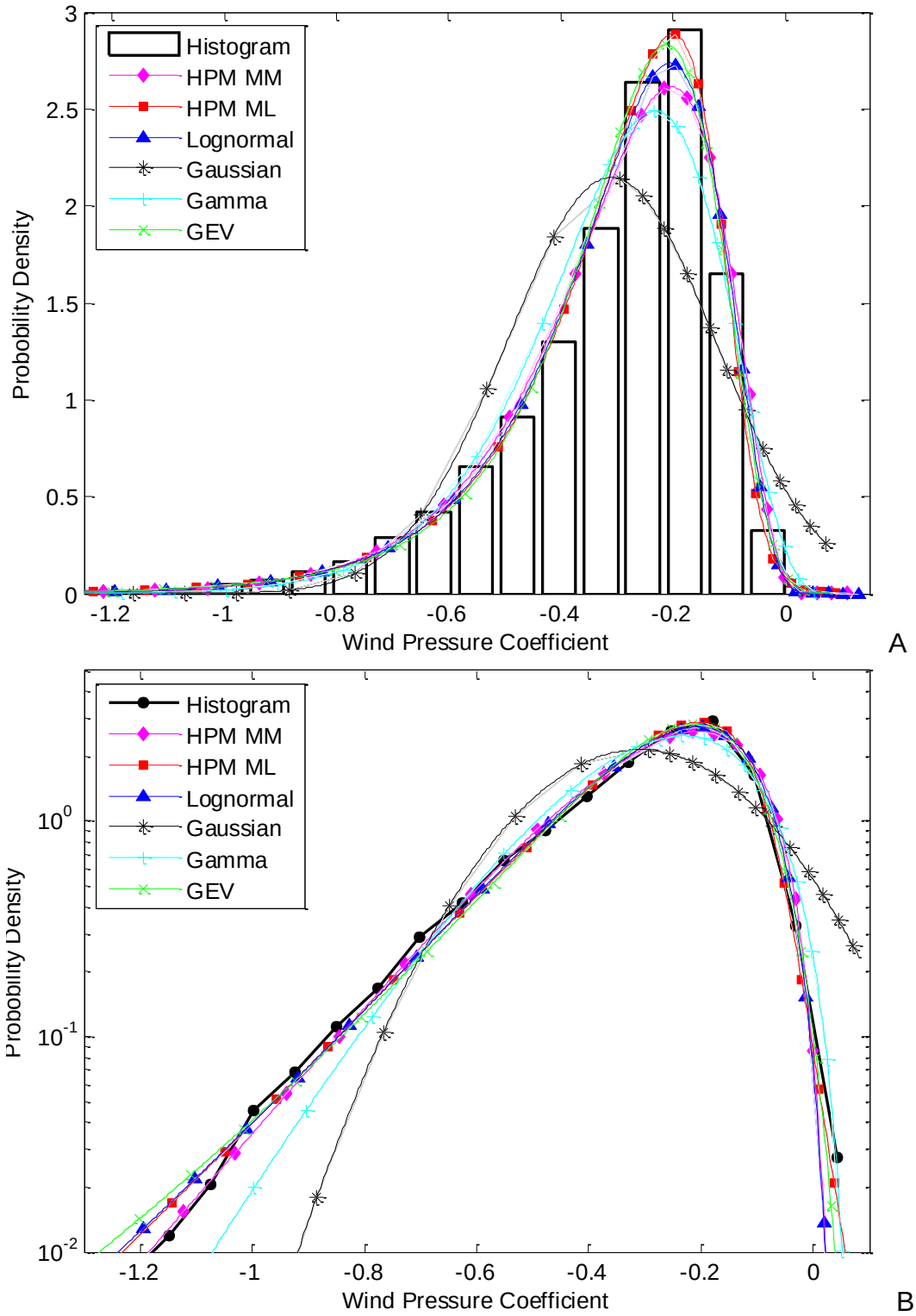


Figure 7-3.Comparison of PDFs of tap 76 (sk=-1.32, kt=5.36) generated by candidate PDF models. A) Linear scale, B) semi-log scale.

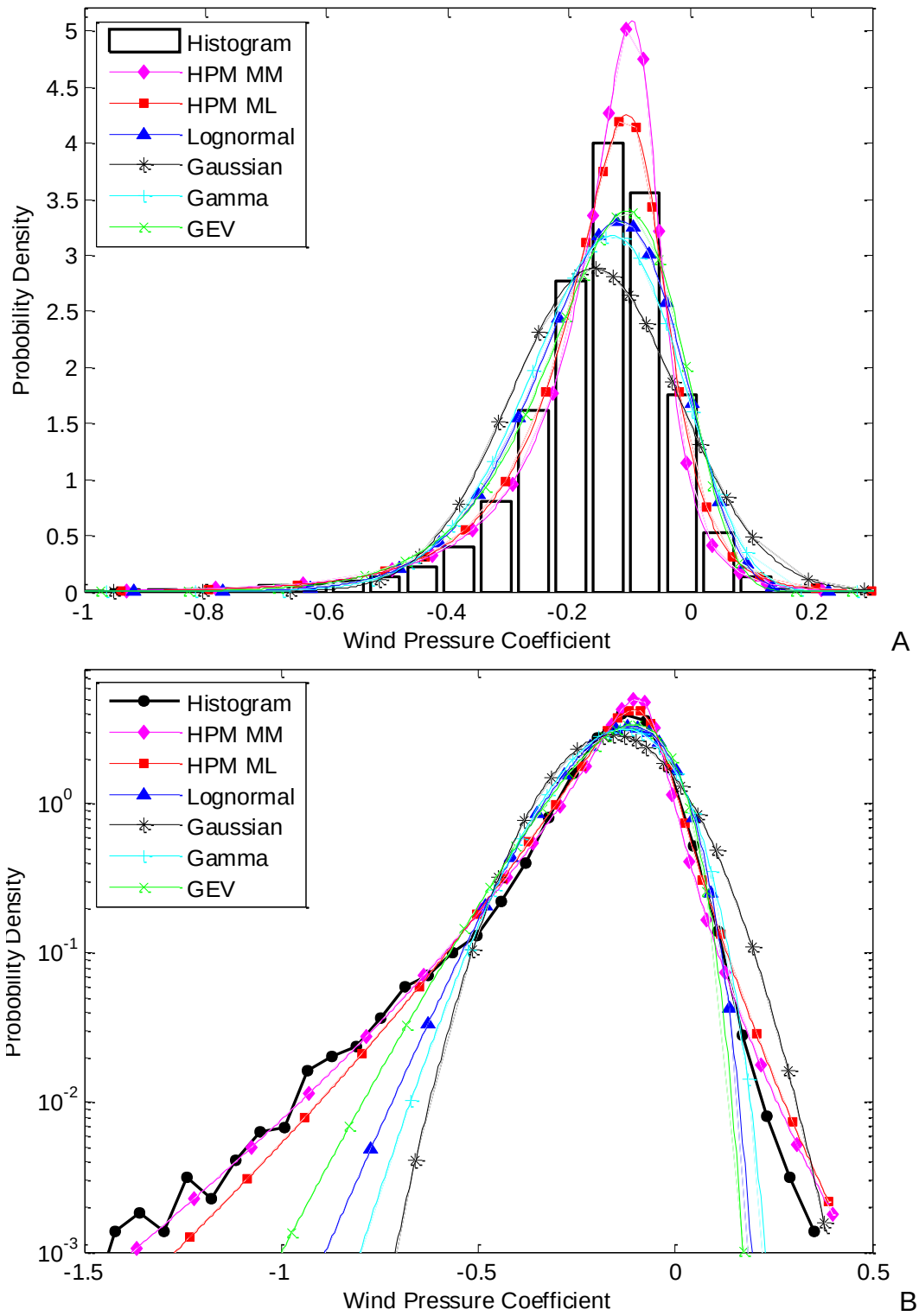


Figure 7-4.Comparison of PDFs of tap 236 (sk=-2.03, kt=12.58) generated by candidate PDF models. A) linear scale, B) semi-log scale.

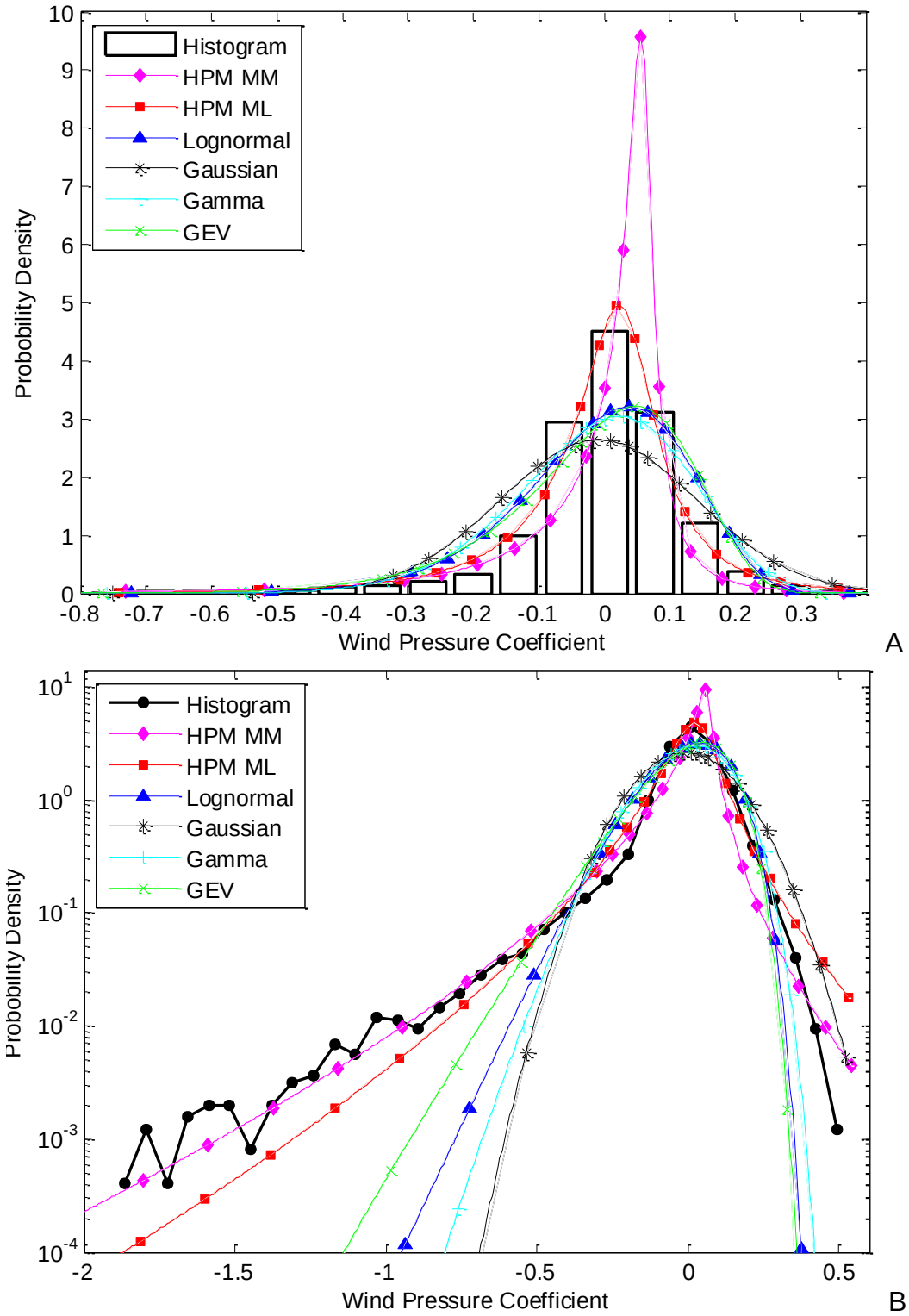


Figure 7-5.Comparison of PDFs of tap 300 (sk=-3.23, kt=25.29) generated by candidate PDF models. A) linear scale, B) semi-log scale.

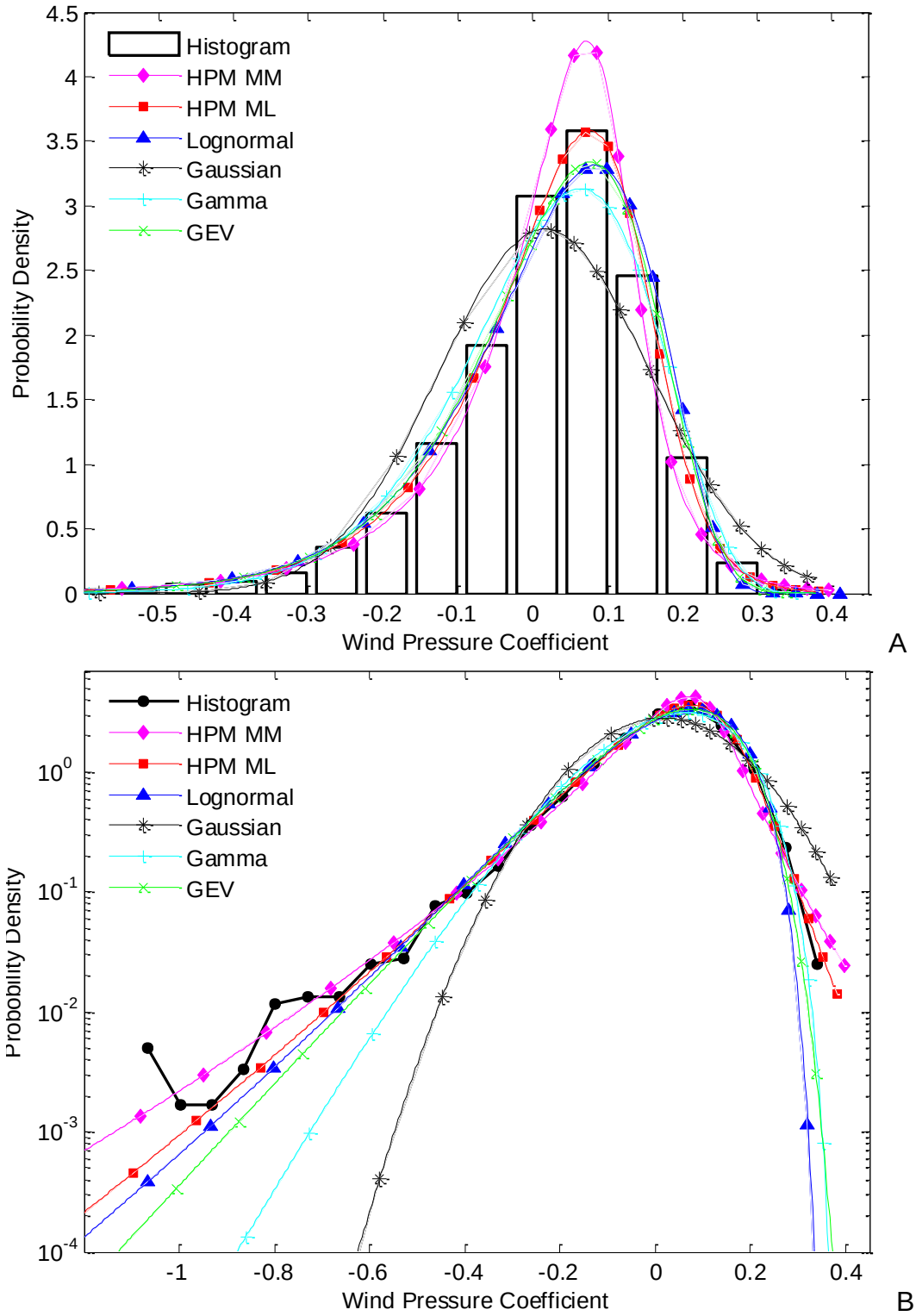


Figure 7-6.Comparison of PDFs of full-scale data Tap 15 (sk=-1.58, kt=9.62) generated by candidate PDF models. A) linear scale, B) semi-log scale.

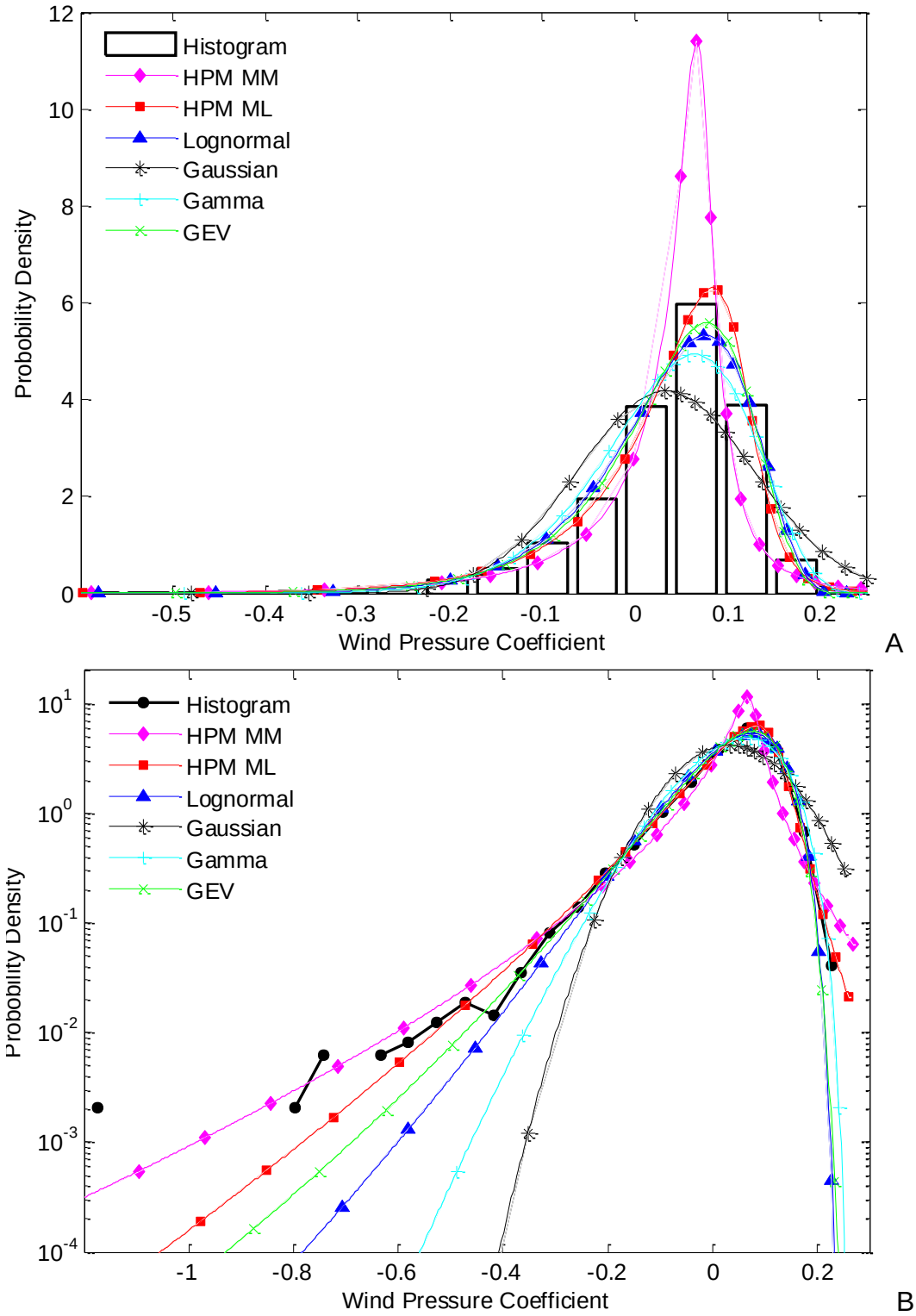
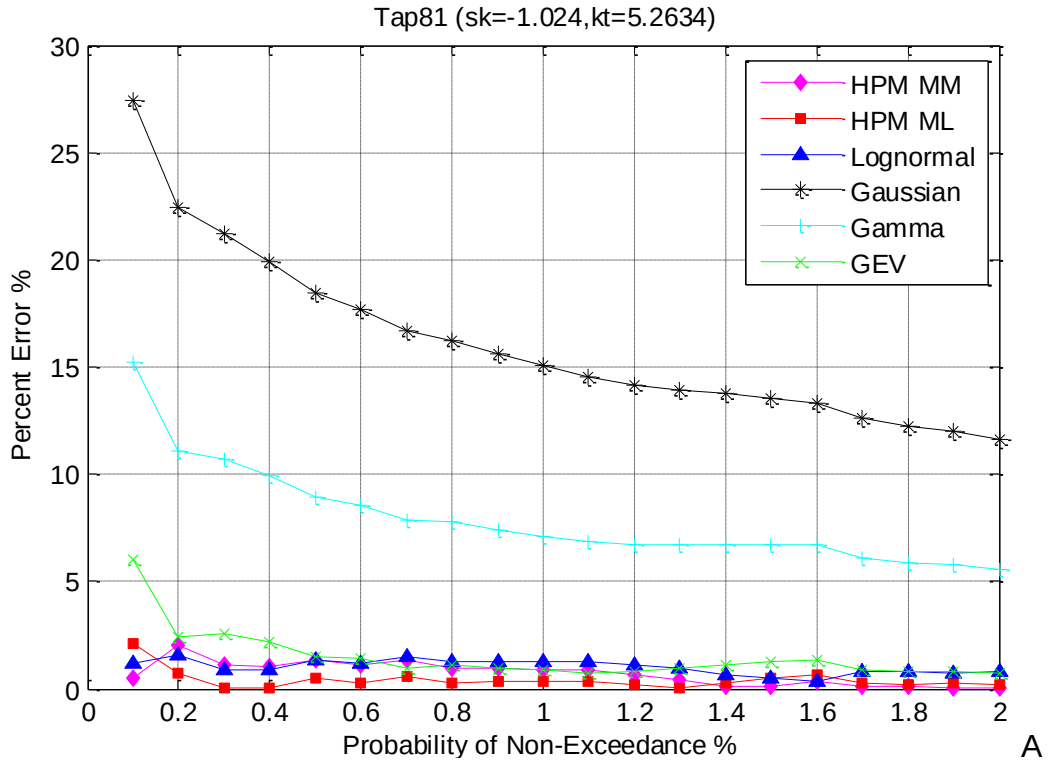
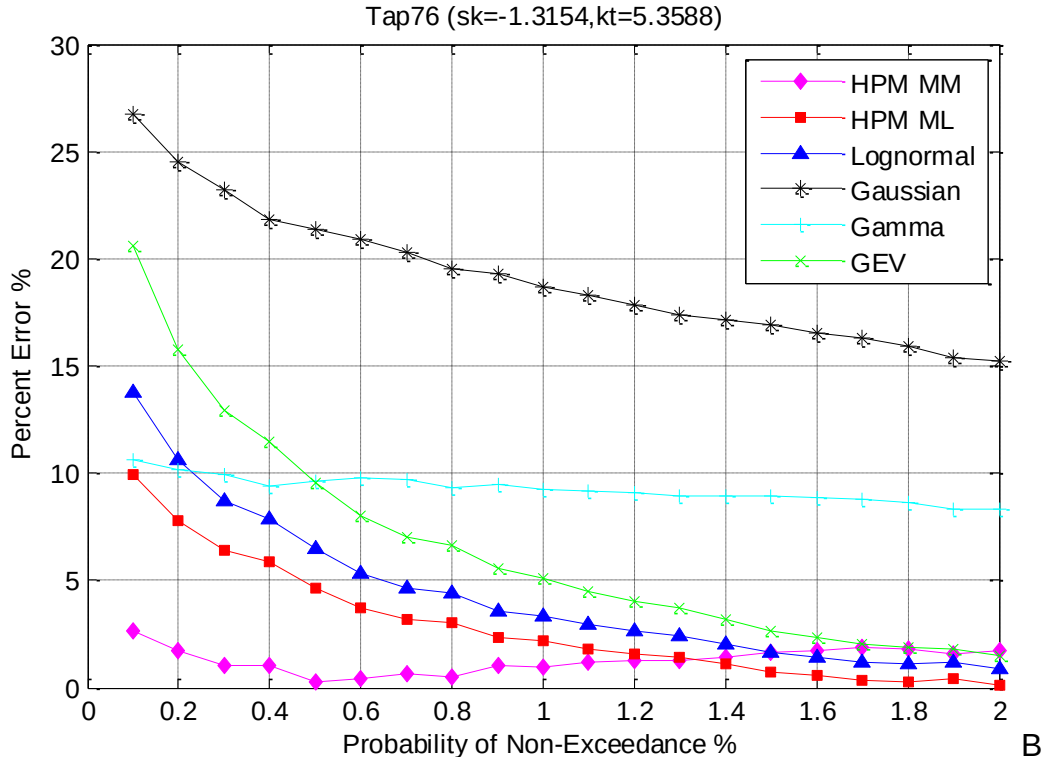


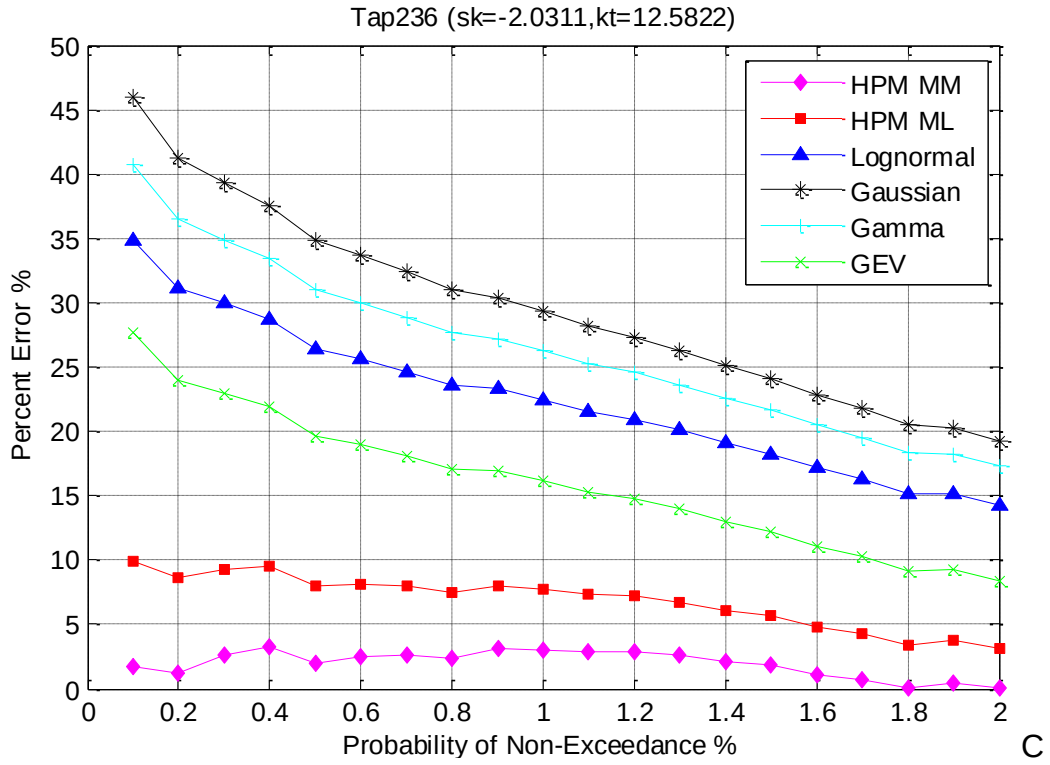
Figure 7-7.Comparison of PDFs of full-scale data Tap24 (sk=-2.59, kt=21.61) generated by candidate PDF models. A) linear scale, B) semi-log scale.


Tap81 (sk=-1.024,kt=5.2634)
Percent Error %
Probability of Non-Exceedance %
HPM MM
HPM ML
Lognormal
Gaussian
Gamma
GEV
0
5
10
15
20
25
30
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2

A


Tap76 (sk=-1.3154,kt=5.3588)
Percent Error %
Probability of Non-Exceedance %
HPM MM
HPM ML
Lognormal
Gaussian
Gamma
GEV
0
5
10
15
20
25
30
0
0.2
0.4
0.6
0.8
1
1.2
1.4
1.6
1.8
2

B



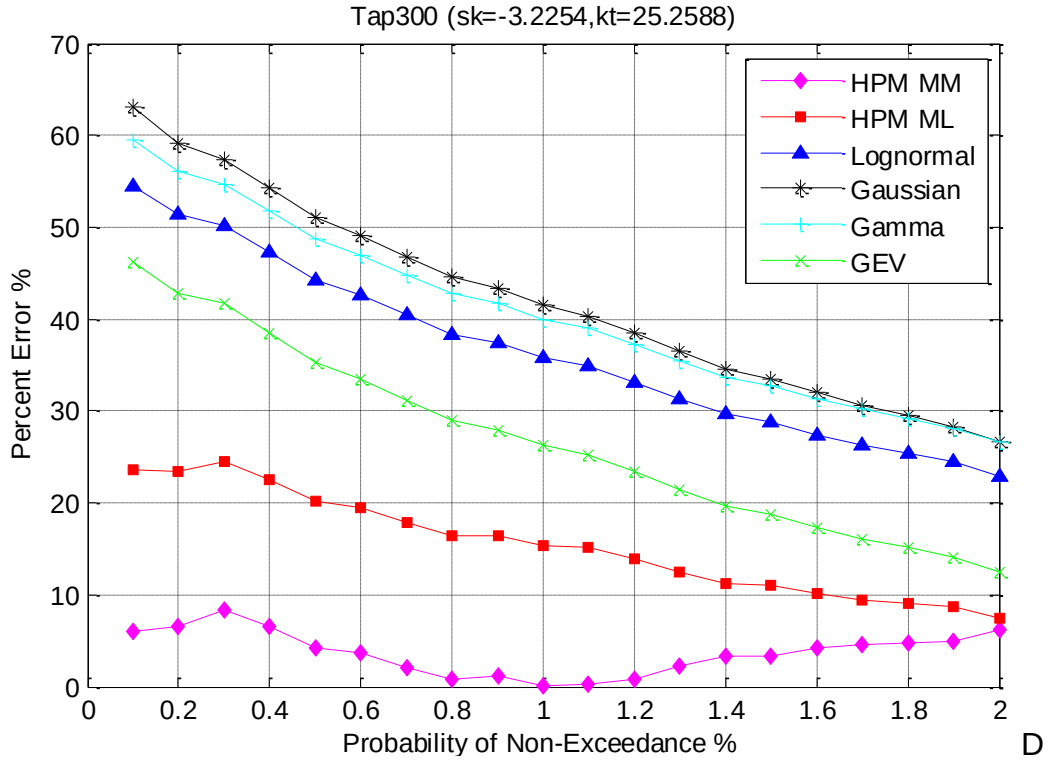


Figure 7-8.Percent error of reference value predicted by PDF models. A) Tap 81, B) Tap 76, C) Tap 236, D) Tap 300.

CHAPTER 8
CONLUSIONS AND RECOMMENDATIONS FOR FUTURE WORK

The research in this thesis has contributed to modeling the PDFs of non-Gaussian data with focus on the tail region, for strongly non-Gaussian data. The work is directly applicable to wind pressure datasets from buildings which exhibit non-Gaussian characteristics. The approaches described have presented a robust method for modeling parent probability density functions of wind pressure in wind engineering applications.

The following sections provide a summary of contributions to and conclusions about the research and present recommendations for future research related to this study area in the thesis. The sections are arranged according to research topics in each chapter.

Summary of Contributions and Conclusions

Chapter 3

Chapter 3 presented the spatial characteristics and statistical properties of wind pressure data occurring on the surface of buildings exposed to natural air flow. It was shown that the wind pressure statistics vary from place to place on the roof and it also varies with wind direction (azimuth). As a result, there can be wide variations in the PDFs of data obtained in pressure taps.

The uncertainty associated with PDF modeling and the statistical variation of wind pressure coefficients were also analyzed. And it was found that the statistical uncertainty embedded in skewness and kurtosis is significant to make the generated PDF deviate from the target PDF, indicating that caution should be taken when trying to use even higher moments, i.e. to 5th or 6th.

Chapter 4

In chapter 4, 5 and 6 three methodologies for estimating the PDF of random data were explored. Chapter 4 describes and application of Pade-Laplace Method to experimental data set. It was found that the PLM generates negative values in the PDF model which is incompatible with a well-behaved PDF.

Chapter 5

Chapter 5 presented a PDF generating approach following the Maximum Entropy Method with traditional and alternative constraints. The results showed that oscillating tail behavior occurs with some models and therefor the Maximum Entropy Method was not sufficiently robust for practical applications.

Chapter 6

In this chapter, the Hermite Polynomial Method is studied, firstly by establishing an effective region where the Hermite Polynomial Method is applicable. This region is defined by mapping the coefficients of the $c_3 - c_4$ points to skewness-kurtosis pairs. Then using this effective region, the HPM was further improved by using surfacing-fitting to develop an approximate closed-form relationship between the data-derived skewness and kurtosis inputs and the model shape parameters.

Finally, the maximum likelihood method was then used to provide an alternative to the method of moments approach for estimating the HPM parameters. Thsis mehod improve the HPM model by reducing the unrealistically large spikes in the mean region of the HPM-generated PDF.

Chapter 7

The distributions of skewness and kurtosis of wind pressure data are compared to the effective regions of the Hermite Polynomial Methods and it was shown that almost

all of the sk-kt pairs lie within the effective region. This is indicative that the HPM is applicable to the wind tunnel and full-scale wind pressure data.

Next, six examples of wind pressure data representing different non-Gaussian effects for wind tunnel data and full-scale data are presented. It is found that: for mild non-Gaussian effects as shown in example 1 and 2, common PDF models could model the PDFs of wind pressure without obvious deviation from the histogram; however, as the non-Gaussian effects became stronger as shown in example 3, 4, 5 and 6, the PDFs generated by common PDF models deviate from the histogram. In all the examples, however, HPM accurately represents the probability content of both weak and strongly non-Gaussian data, whereas the more common models have mixed performance, and provide poor results for the strongly non-Gaussian data. The HPM show excellent performance in the tail region even for the strongly non-Gaussian case.

Also, HPM, either based on method of moment or maximum likelihood method, demonstrates better performance in predicting the reference values in the tail region than common PDF models.

Recommendations for Future Work

Although in this study the HPM is only applied to one particular wind tunnel dataset and one full-scale dataset, the effective region and the distribution of sk-kt pairs of the two datasets indicate that HPM has high potential to serve as a flexible and robust PDF model for wind pressure data. Further study of the applicability of HPM is ongoing with additional full-scale and scale model bluff body pressure datasets.

For design and structure reliability purposes, it is necessary to estimate the largest value (extreme value) of the wind pressure and internal forces. There are mainly two classes of methods to estimate the extreme value. One class of methods is fitting an

exponential distribution of a population consisting only of large amplitude ‘spikes’ in the data record (e.g. (Peterka 1983), (Simiu and Heckert 1996)). Since this class of analysis uses very limited information in the data record, the large spikes, the estimated extreme value has large variability.

The other category of methods introduced by (Sadek and Simiu 2002) uses all the data contained in the time series. This method provides more stable estimates and yields useful information on the probability distribution of the peaks. The estimation of non-Gaussian process is based on the standard translation processes approach. The first step for the estimation requires the identification and evaluation of the optimal parent probability distribution. (Sadek and Simiu 2002) and (Tieleman, Ge and Hajj 2007) showed that the distribution of the time series of internal force or wind pressure coefficients is well represented by 3-parameter gamma distribution. However, Sadek and Simiu also pointed out that gamma distribution is not an accurate probabilistic reprentation of the distribution of non-Gaussian time series: the gamma distibution is appropriate for representing the longer tail; it is not appropriate for representing the shorter tail. To overcome the problem, they used gamma distribution to estimate maximum peran, while a normal distribution was applied to estimate the minimum peak.

In this study, it is also found that, even for the longer tail, the gamma distribution deviates dramatically from the histogram of datasets have strong non-Gaussian effects. Thus, since the HPM is flexible and, more importantly, much more precise in the tail regions, it is interesting to investigate whether the estimated extreme value of wind pressure or structural loads will be improved if HPM is used as an alternative parent

PDF model. A follow up study will consider the implications of the HPM with regard to expected extreme value pressures and area averaged pressures.

APPENDIX A
CONTOUR PLOTS FOR AZIMUTH 135^0 AND 180^0

This appendix contains the contour plots for azimuth 135^0 and 180^0. For these two circumstances, the spatial variation of the statistical properties of wind pressure coefficients is largely similar to 45^0 and 0^0, respectively, except that the azimuth are different.

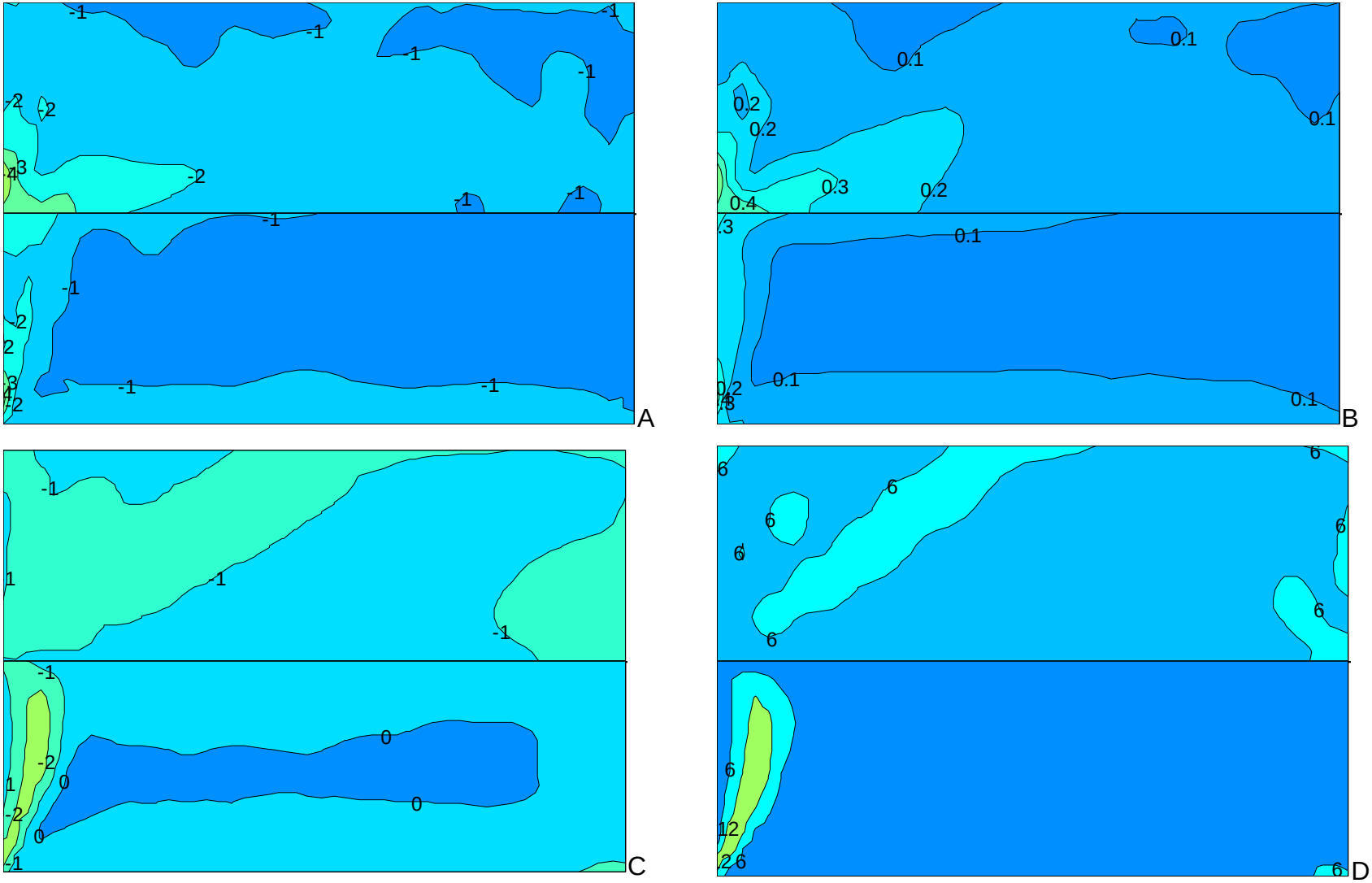


Figure A-1.Spatial variation of statistical properties of Cps for azimuth 135^0. A) peak value, B) standard deviation, C) skewness, D) kurtosis.

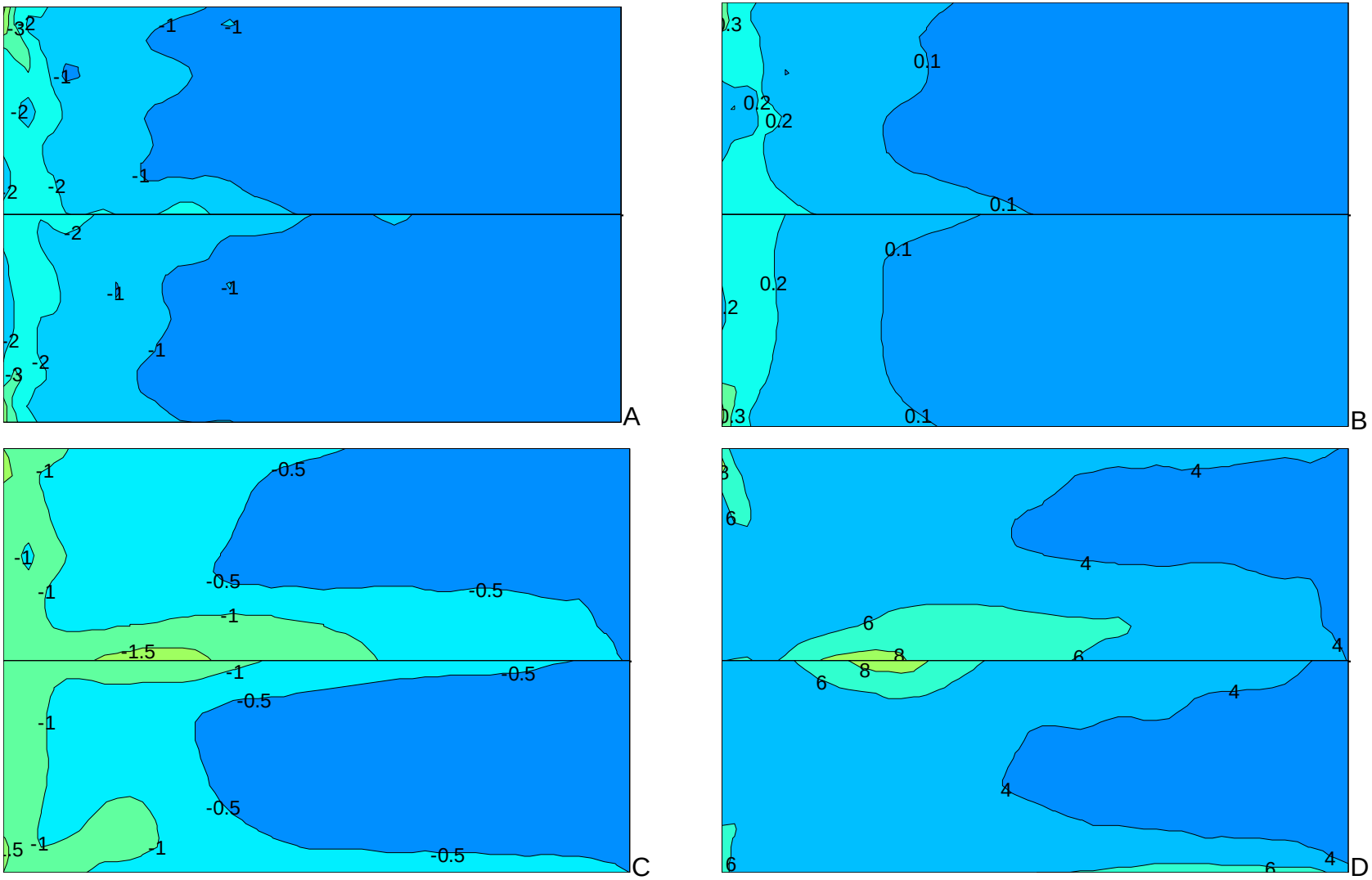


Figure A-2.Spatial variation of statistical properties of Cps for azimuth180^0 . A) peak value, B) standard deviation, C) skewness, D) kurtosis.

APPENDIX B
PDFS GENERATED BY HPM FOR VARIOUS ROOF REGIONS

In Chapter 7, six examples are presented for conciseness. Because the cornering winds generates the most severe non-Gaussian effect among other wind directions, only the circumstance of azimuth=45^0 will be presented. Appendix B presents results which give people a better idea of how well HPM (including MM and ML) works for taps throughout the roof, that is the flexibilty of HPM. 16 taps are evenly chosen as representative of total 387 taps as shown in Figure B-1.

By observing the figures from Figure B-2 to Figure B-5, PDFs generated HPM track the data histogram very well for all the taps chosen. In this process, there is no need to change PDF model from particular roof region to another; HPM works well for all the region on the roof, even though the shapes of data histogram change greatly (a wide range of skewness and kurtosis). This demonstrates the flexibility and robustness of HPM.

Observing PDFs of the 16 taps, there is not significant difference between PDFs generated by HPM-MM and PDFs generated by HPM-ML. However, when the magnitude of skewness and kurtosis is large, HPM-MM will generate high ‘spike’; for these cases, HPM-ML provides a better fit in the mean region, effectively reducing the high ‘spark’ of HPM-MM in the mean region.

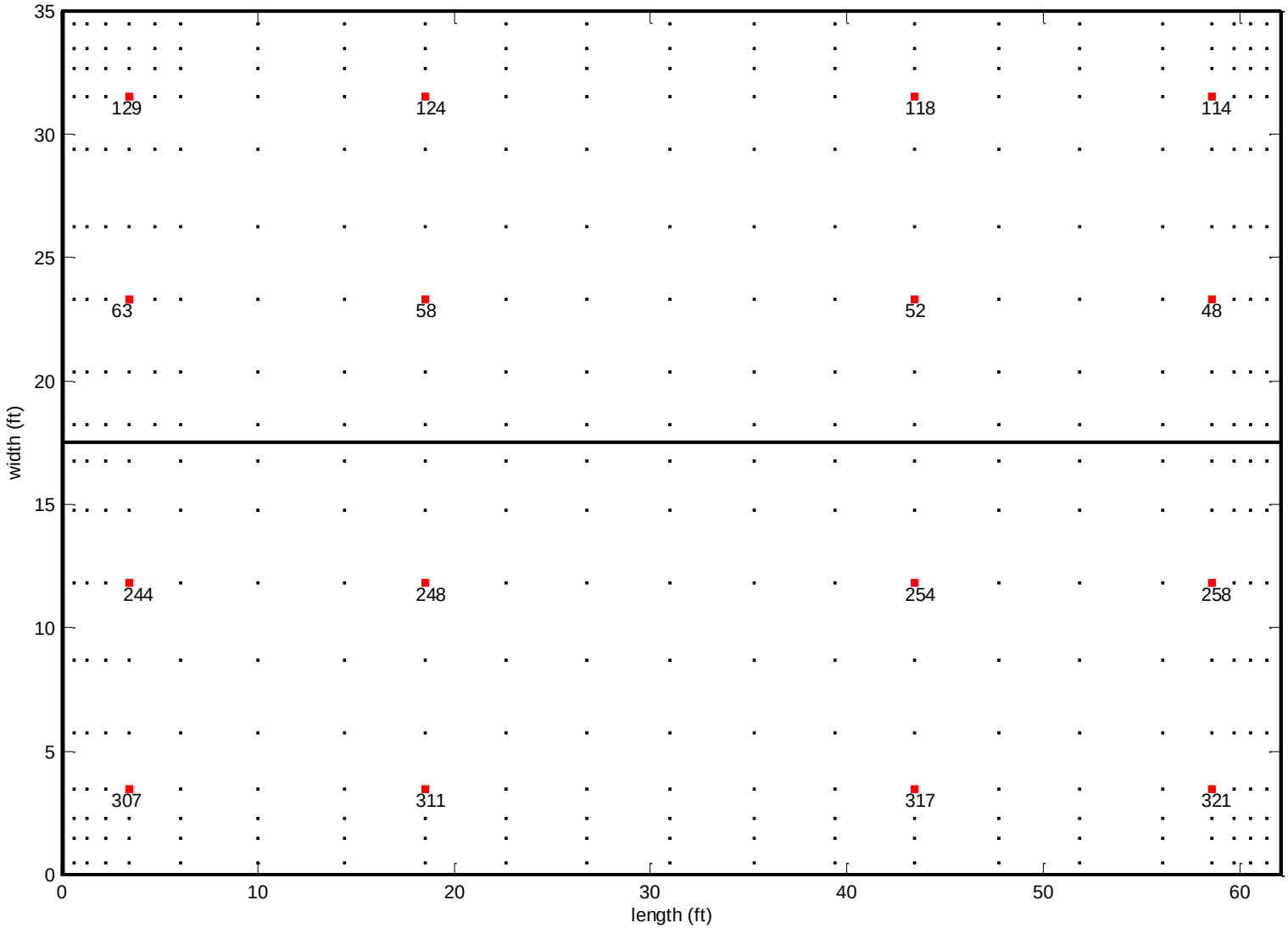


Figure B-1. Layout of chosen taps for modeling of PDFs.

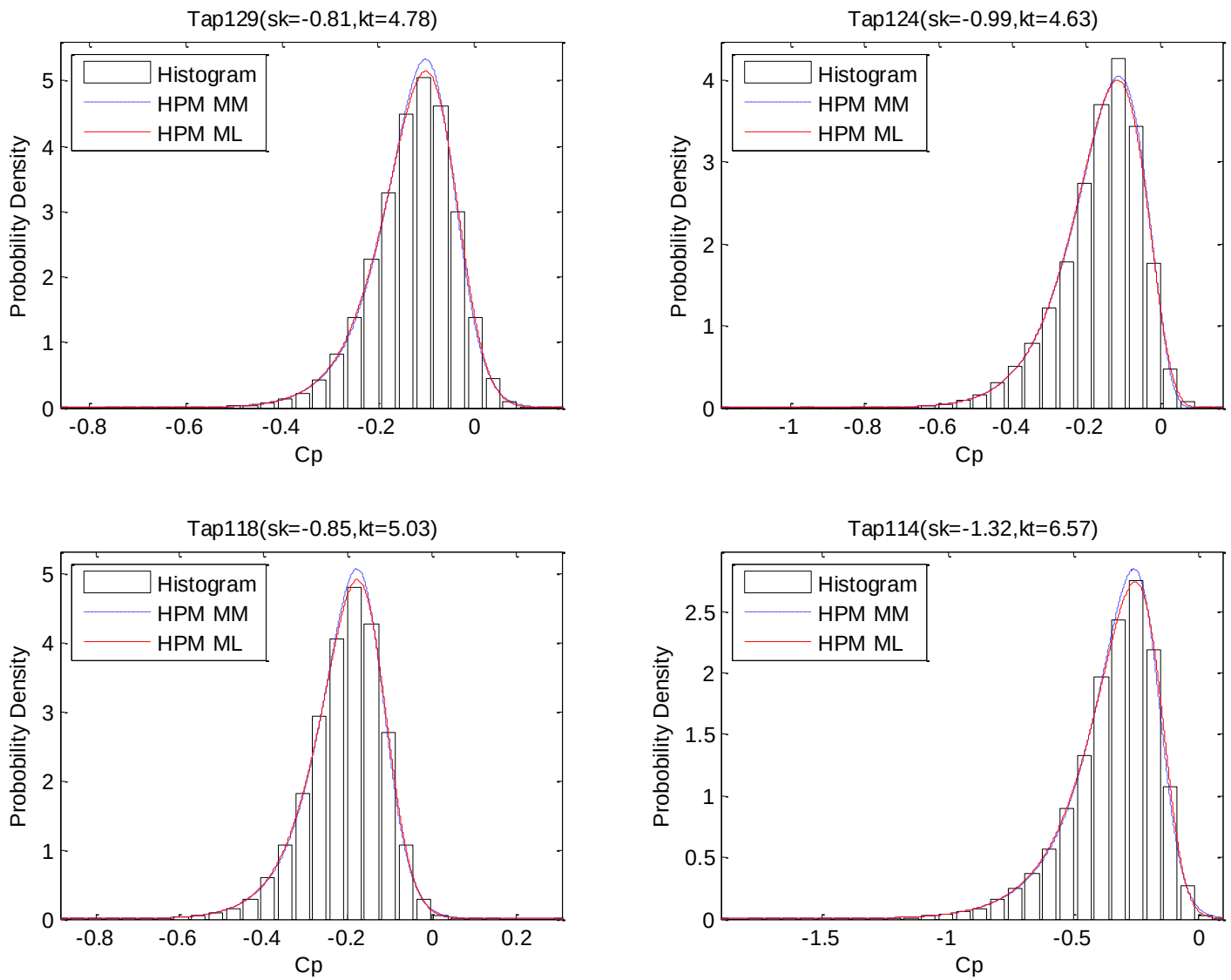


Figure B-2. PDFs generated by HPM for the first row of taps.

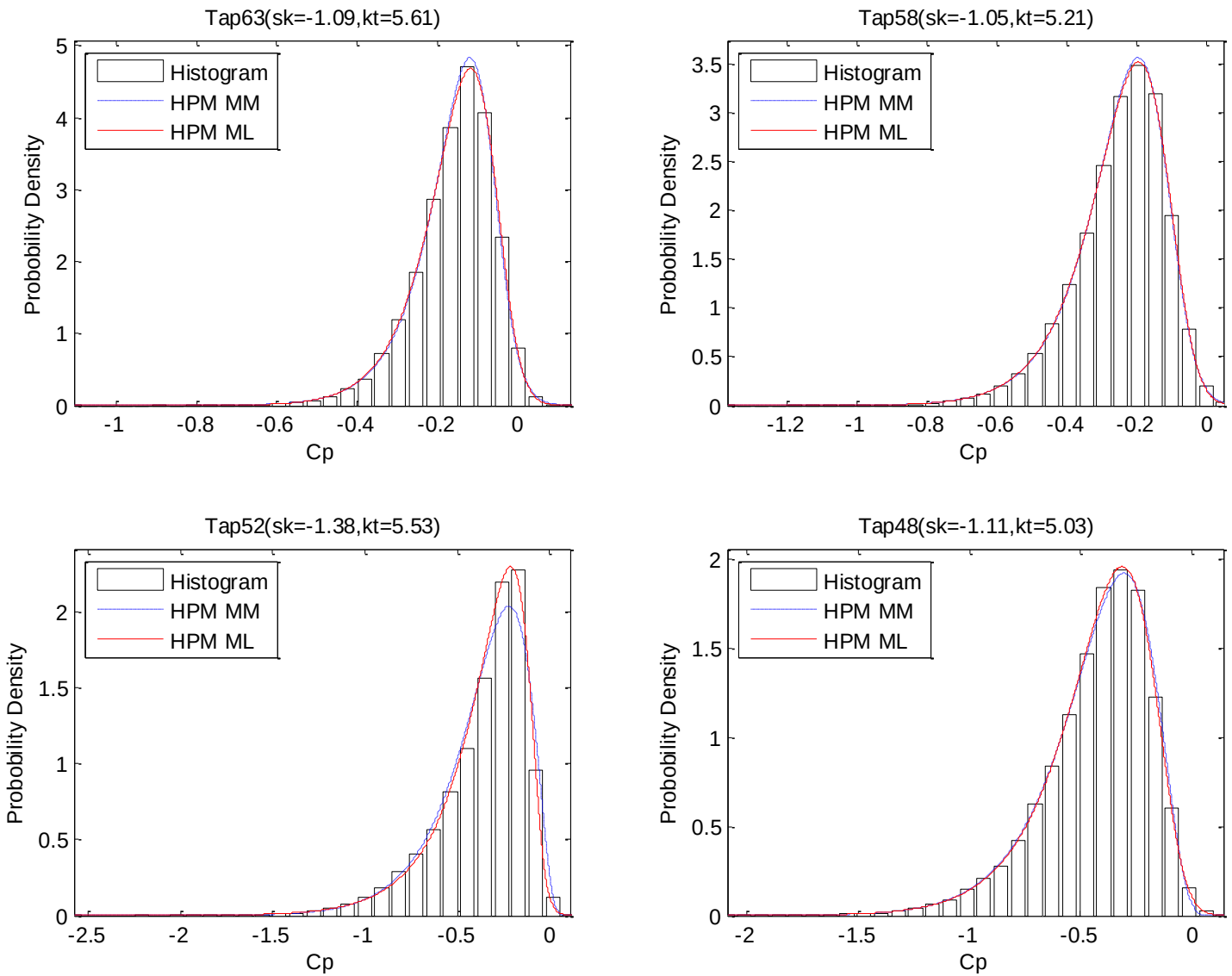


Figure B-3. PDFs generated by HPM for the second row of taps.

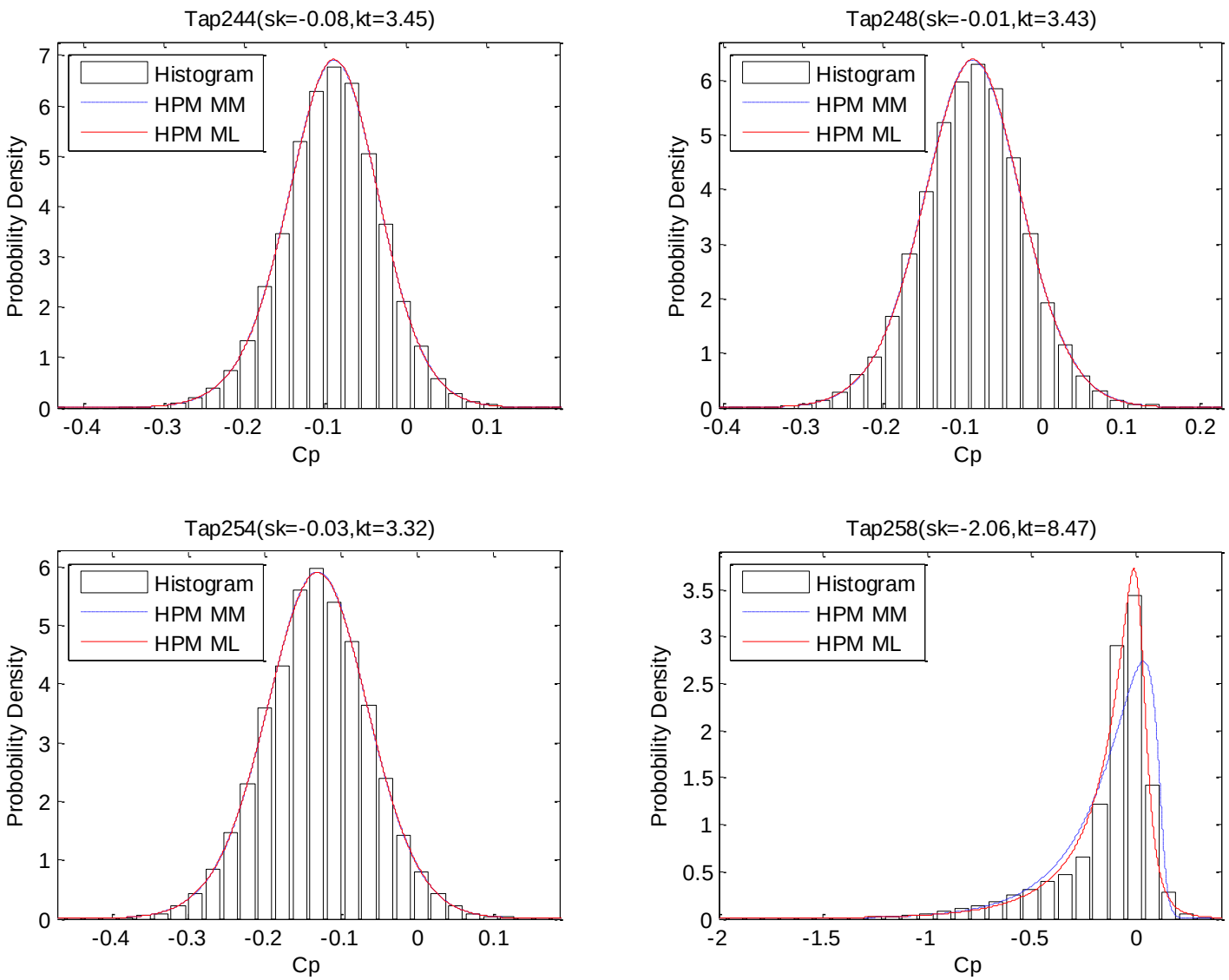


Figure B-4. PDFs generated by HPM for the third row of taps.

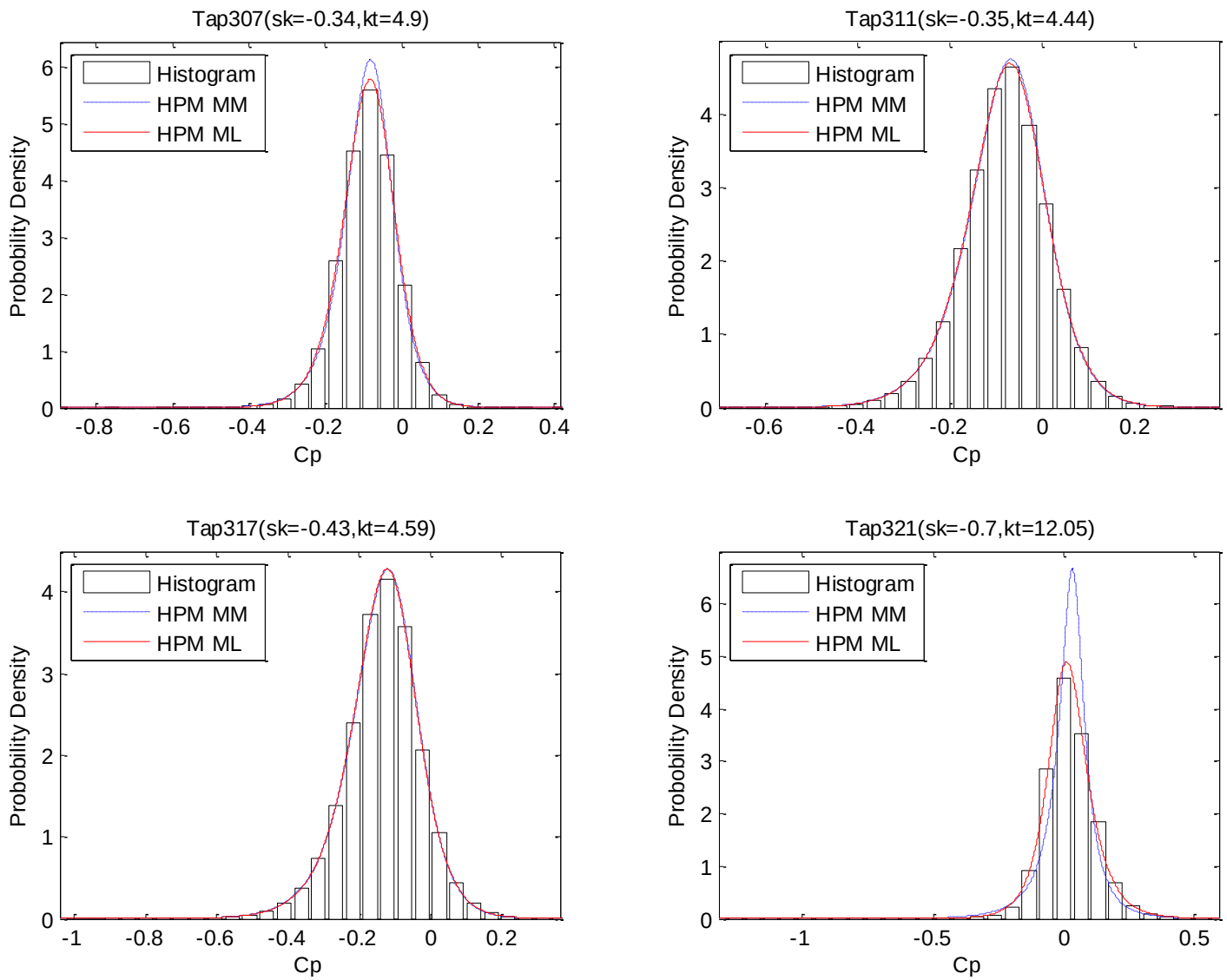


Figure B-5. PDFs generated by HPM for the fourth row of taps.

LIST OF REFERENCES

Agmon, N., Y. Alhassid & R. D. Levine (1979) An algorithm for finding the distribution of maximal entropy. *Journal of Computational Physics,* 30**,** 250-258.

Alhassid, Y., N. Agmon & R. D. Levine (1978) An upper bound for the entropy and its applications to the maximal entropy problem. *Chemical Physics Letters,* 53**,** 22-26.

Amindavar, H. & J. A. Ritcey (1994) Pade approximations of probability density functions. *IEEE Transactions on Aerospace and Electronic Systems,* 30**,** 416-424.

Anderson, T. W. & D. A. Darling (1954) A Test of Goodness of Fit. *Journal of the American Statistical Association,* 49**,** 765-769.

Atkinson, K. E. 1989. *An introduction to numerical analysis / Kendall E. Atkinson*. New York: Wiley.

Aubard, J., P. Levoir, A. Denis & P. Claverie (1987) DIRECT ANALYSIS OF CHEMICAL RELAXATION SIGNALS BY A METHOD BASED ON THE COMBINATION OF LAPLACE TRANSFORM AND PADE APPROXIMANTS. *Computers and Chemistry,* 11**,** 163-178.

Bajzer, Z., A. C. Myers, S. S. Sedarous & F. G. Prendergast (1989) Padé-Laplace method for analysis of fluorescence intensity decay. *Biophysical Journal,* 56**,** 79-93.

Chatfield, C. (1995) Model Uncertainty, Data Mining and Statistical Inference. *Journal of the Royal Statistical Society. Series A (Statistics in Society),* 158**,** 419-466.

Cope, A. D., K. R. Gurley, M. Gioffre & T. A. Reinhold (2005) Low-rise Gable Roof Wind Loads: Characterization and Stochastic Simulation. *Journal of Wind Engineering and Industrial Aerodynamics,* 93**,** 719-738.

Datin, P. L. & D. O. Prevatt. 2007. Wind Uplift Reactions at Roof-to-Wall Connections of Wood-Framed Gable Roof Assembly. 12th International Conference on Wind Engineering: Australasian Wind Engineering Society, Cairns, Australia.

Davenport, A. G. (1961) Application of statistical concepts to wind loading of structures. *Institution of Civil Engineers -- Proceedings,* 20**,** 449-472.

Draper, D. (1995) Assessment and Propagation of Model Uncertainty. *Journal of the Royal Statistical Society. Series B (Methodological),* 57**,** 45-97.

Ginger, J. D. & C. W. Letchford (1993) Characteristics of large pressures in regions of flow separation. *Journal of Wind Engineering and Industrial Aerodynamics,* 49**,** 301-310.

Grigoriu, M. (1984) CROSSINGS OF NON-GAUSSIAN TRANSLATION PROCESSES. *Journal of Engineering Mechanics,* 110**,** 610-620.

Holmes, J. D. (1981) NON-GAUSSIAN CHARACTERISTICS OF WIND PRESSURE FLUCTUATIONS. *Journal of Wind Engineering and Industrial Aerodynamics,* 7**,** 103-108.

Holmes, J. D. & L. S. Cochran (2003) Probability distributions of extreme pressure coefficients. *Journal of Wind Engineering and Industrial Aerodynamics,* 91**,** 893-901.

Holmes, J. D. A. 2001. *Wind Loading of Structures [electronic resource]*. New York : Florence: Spon Press [Imprint] Routledge Taylor & Francis Group [Distributor.

Jay, E., J. P. Ovarlez & P. Duvaut (2000) New methods of radar performances analysis. *Signal Processing,* 80**,** 2527-2540.

Jaynes, E. T. (1957) Information theory and statistical mechanics. *Physical Review,* 106**,** 620-630.

Kareem & A. 1978. Wind Excited Motion of Tall Buildings. Fort Collins, CO.: Dissertation in partial fulfillment of the degree of doctor of philosophy, Colorado State University.

Kumar, K. (1992) Padé approximation and its application in time series analysis. *Applied Mathematics and Computation,* 48**,** 139-151.

Kumar, K. S. & T. Stathopoulos (1998) Fatigue analysis of roof cladding under simulated wind loading. *Journal of Wind Engineering and Industrial Aerodynamics,* 77-78**,** 171-183.

Lee, M. & C.-M. Ho (1990) Lift force of delta wings. *Applied Mechanics Reviews,* 43**,** 209-209.

Li, Q. S., I. Calderone & W. H. Melbourne (1999) Probabilistic characteristics of pressure fluctuations in separated and reattaching flows for various free-stream turbulence. *Journal of Wind Engineering and Industrial Aerodynamics,* 82**,** 125-145.

Liu, Z., D. O. Prevatt, L. D. Aponte-Bermudez, K. R. Gurley, T. A. Reinhold & R. E. Akins (2009) Field measurement and wind tunnel simulation of hurricane wind loads on a single family dwelling. *Engineering Structures,* 31**,** 2265-2274.

Massey, F. J., Jr. (1951) The Kolmogorov-Smirnov Test for Goodness of Fit. *Journal of the American Statistical Association,* 46**,** 68-78.

Mensah, A. F. 2010. *Determination of wind uplift forces using Database-Assisted Design (DAD) approach for light framed wood structures [electronic resource] / by Akwasi Frimpong Mensah*. [Gainesville, Fla: University of Florida.

Mensah, A. F., P. L. Datin, D. O. Prevatt, R. Gupta & J. W. van de Lindt (2011) Database-assisted design methodology to predict wind-induced structural behavior of a light-framed wood building. *Engineering Structures,* 33**,** 674-684.

Ochi, M. K. (1986) Non-Gaussian random processes in ocean engineering. *Probabilistic Engineering Mechanics,* 1**,** 28-39.

Ormoneit, D. & H. White (1999) An efficient algorithm to compute maximum entropy densities. *Econometric Reviews,* 18**,** 127-140.

Peterka, J. A. (1983) SELECTION OF LOCAL PEAK PRESSURE COEFFICIENTS FOR WIND TUNNEL STUDIES OF BUILDINGS. *Journal of Wind Engineering and Industrial Aerodynamics,* 13**,** 477-488.

Reed, D. A. (1993) Influence of non-Gaussian local pressures on cladding glass. *Journal of Wind Engineering and Industrial Aerodynamics,* 48**,** 51-61.

Rockinger, M. & E. Jondeau (2002) Entropy densities with an application to autoregressive conditional skewness and kurtosis. *Journal of Econometrics,* 106**,** 119-142.

Sadek, F. & E. Simiu (2002) Peak Non-Gaussian Wind Effects for Database-assisted Low-rise Building Design. *Journal of Engineering Mechanics,* 128**,** 530-539.

Simiu, E. & N. A. Heckert (1996) Extreme wind distribution tails: a 'peaks over threshold' approach. *Journal of structural engineering New York, N.Y.,* 122**,** 539-547.

Simiu, E. & T. Miyata. 2006. *Design of Buildings and Bridges for Wind, A Practical Guide for ASCE-7 Standard Users and Designers of Special Structures*. New Jersey: John Wiley & Sons.

Sobczyk, K. & J. Trebicki (1993) Maximum entropy principle and nonlinear stochastic oscillators. *Physica A: Statistical Mechanics and its Applications,* 193**,** 448-468.

Stathopoulos, T. (1980) PDF OF WIND PRESSURES ON LOW-RISE BUILDINGS. 106**,** 973-990.

Tieleman, H. W., Z. Ge & M. R. Hajj (2007) Theoretically estimated peak wind loads. *Journal of Wind Engineering and Industrial Aerodynamics,* 95**,** 113-132.

Tognarelli, M. A., J. Zhao & A. Kareem (1997) Equivalent Statistical Cubicization for System and Forcing Nonlinearities. *Journal of Engineering Mechanics,* 123**,** 890-893.

Winterstein, S. R. 1987. MOMENT-BASED HERMITE MODELS OF RANDOM VIBRATION. var paging-var paging.

Winterstein, S. R. (1988) NONLINEAR VIBRATION MODELS FOR EXTREMES AND FATIGUE. *Journal of Engineering Mechanics,* 114**,** 1772-1790.

Winterstein, S. R. & T. Kashef (2000) Moment-based Load and Response Models with Wind Engineering Application. *Journal of Solar Energy Engineering, Transactions of the ASME,* 122**,** 122-128.

Wu, X. (2003) Calculation of maximum entropy densities with application to income distribution. *Journal of Econometrics,* 115**,** 347-354.

Zellner, A. & R. A. Highfield (1988) Calculation of maximum entropy distributions and approximation of marginalposterior distributions. *Journal of Econometrics,* 37**,** 195-209.

BIOGRAPHICAL SKETCH

Luping Yang was born in Sichuan Province, China. He obtained his Bachelor of Engineering Degree from Department of Applied Mechanics and Aerospace Engineering in Tongji University in Shanghai, China, in July 2009. In the summer of 2008, he also served as a research assistant in Institute of Mechanics, Chinese Academy of Science, where he worked in National Micro-Gravity Laboratory and learned about the technique of Particle Image Velocimetry (PIV). After graduation, he joined He joined University of Florida in pursuit of a Master of Science degree in August 2009. He received a degree of Master of Science in civil engineering and a minor in mathematics in the fall of 2011.

Luping Yang is a student member of the American Association for Wind Engineering and the American Society of Civil Engineers.